ENCYCLOPEDIA OF SOILS IN THE ENVIRONMENT

EDITED BY
DANIEL HILLEL
CYNTHIA ROSENZWEIG
DAVID POWLSON
KATE SCOW
MICHAIL SINGER
DONALD SPARKS

VOLUME TWO

ENCYCLOPEDIA OF SOILS IN THE ENVIRONMENT

FOUR-VOLUME SET

by Daniel Hillel (Editor-in-Chief)

Hardcover: 2200 pages
Publisher: Academic Press; 1 edition (November 8, 2004)
Language: English
ISBN-10: 0123485304
ISBN-13: 978-0123485304

Book Description

More than ever before, a compelling need exists for an encyclopedic resource about soil the rich mix of mineral particles, organic matter, gases, and soluble compounds that foster both plant and animal growth. Civilization depends more on the soil as human populations continue to grow and increasing demands are placed upon available resources.
The Encyclopedia of Soils in the Environmentis a comprehensive and integrated consideration of a topic of vital importance to human societies in the past, present, and future.

This important work encompasses the present knowledge of the world's variegated soils, their origins, properties, classification, and roles in the biosphere. A team of outstanding, international contributors has written over 250 entries that cover a broad range of issues facing today's soil scientists, ecologists, and environmental scientists.
This four-volume set features thorough articles that survey specific aspects of soil biology, ecology, chemistry and physics. Rounding out the encyclopedia's excellent coverage, contributions cover cross-disciplinary subjects, such as the history of soil utilization for agricultural and engineering purposes and soils in relation to the remediation of pollution and the mitigation of global climate change.

This comprehensive, yet accessible source is a valuable addition to the library of scientists, researchers, students, and policy makers involved in soil science, ecology, and environmental science.

Also available online via ScienceDirect featuring extensive browsing, searching, and internal cross-referencing between articles in the work, plus dynamic linking to journal articles and abstract databases, making navigation flexible and easy. For more information, pricing options and availability visit www.info.sciencedirect.com.

* A distinguished international group of editors and contributors
* Well-organized encyclopedic format providing concise, readable entries, easy searches, and thorough cross-references
* Abundant visual resources — photographs, figures, tables, and graphs — in every entry
* Complete up-to-date coverage of many important topics — essential information for scientists, students and professionals alike

EDITOR-IN-CHIEF

Daniel Hillel
Columbia University
New York, NY
USA

EDITORS

Jerry L Hatfield
National Soil Tilth Laboratory
Ames, IA
USA

David S Powlson
Rothamsted Research
Harpenden
UK

Cynthia Rosenzweig
NASA Goddard Institute for Space Studies
New York, NY
USA

Kate M Scow
University of California
Davis, CA
USA

Michael J Singer
University of California
Davis, CA
USA

Donald L Sparks
University of Delaware
Newark, DE
USA

FOREWORD

The *Encyclopedia of Soils in the Environment* is a vitally important scientific publication and an equally important contribution to global public policy. The *Encyclopedia* brings together a remarkable range of cutting-edge scientific knowledge on all aspects of soil science, as well as the links of soils and soil science to environmental management, food production, biodiversity, climate change, and many other areas of significant concern. Even more than that, the *Encyclopedia* will immediately become an indispensable resource for policy makers, analysts, and students who are focusing on one of the greatest challenges of the 21st century. With 6.3 billion people, our planet is already straining to feed the world's population, and is failing to do so reliably in many parts of the world. The numbers of chronically poor in the world have been stuck at some 800 million in recent years, despite long-standing international goals and commitments to reduce that number by several hundred million. Yet the challenge of food production will intensify in coming decades, as the human population is projected to rise to around 9 billion by mid-century, with the increased population concentrated in parts of the world already suffering from widespread chronic under-nourishment.

Unless the best science is brought to these problems, the situation is likely to deteriorate sharply. Food production systems are already under stress, for reasons often related directly to soils management. In Africa, crop yields are disastrously low and falling in many places due to the rampant depletion of soil nutrients. This situation needs urgent reversal, through increasing use of agro-forestry techniques (e.g. inter-cropping cereals with leguminous nitrogen-fixing trees) and increasing the efficient applications of chemical fertilizers. In other impoverished, as well as rich, parts of the planet, decades of intensive agriculture under irrigation have led to salinization, water-logging, eutrophication of major water bodies, dangerous declines of biodiversity and other forms of environmental degradation. These enormous strains are coupled with the continuing pressures of tropical deforestation and the lack of new promising regions for expanding crop cultivation to meet the needs of growing populations. Finally, there looms the prospect of anthropogenic climate change. Global warming and associated complex and poorly understood shifts in precipitation extremes and other climate variables all threaten the world's natural ecosystems and food production systems in profound yet still imperfectly understood ways. The risks of gradual or abrupt climate change are coupled with the risks of drastic perturbations to regional and global food supplies.

The *Encyclopedia* offers state-of-the-art contributions on each of these challenges, as well as links to entries on the fundamental biophysical processes that underpin the relevant phenomena. The world-scale and world-class collaboration that stands behind this unique project signifies its importance for the world community. It is an honor and privilege for me to introduce this path-breaking endeavor.

Jeffrey D Sachs
Director
The Earth Institute at Columbia University
Quetelet Professor of Sustainable Development
Columbia University, New York, USA

PREFACE

The term 'soil' refers to the weathered and fragmented outer layer of our planet's land surfaces. Formed initially through the physical disintegration and chemical alteration of rocks and minerals by physical and biogeochemical processes, soil is influenced by the activity and accumulated residues of a myriad of diverse forms of life. As it occurs in different geologic and climatic domains, soil is an exceedingly variegated body with a wide range of attributes.

Considering the height of the atmosphere, the thickness of the earth's rock mantle, and the depth of the ocean, one observes that soil is an amazingly thin body – typically not much more than one meter thick and often less than that. Yet it is the crucible of terrestrial life, within which biological productivity is generated and sustained. It acts like a composite living entity, a home to a community of innumerable microscopic and macroscopic plants and animals. A mere fistful of soil typically contains billions of microorganisms, which perform vital interactive biochemical functions. Another intrinsic attribute of the soil is its sponge-like porosity and its enormous internal surface area. That same fistful of soil may actually consist of several hectares of active surface, upon which physicochemical processes take place continuously.

Realizing humanity's utter dependence on the soil, ancient peoples, who lived in greater intimacy with nature than many of us today, actually revered the soil. It was not only their source of livelihood, but also the material from which they built their homes and that they learned to shape, heat, and fuse into household vessels and writing tablets (ceramic, made of clayey soil, being the first synthetic material in the history of technology). In the Bible, the name assigned to the first human was Adam, derived from 'adama,' meaning soil. The name given to that first earthling's mate was Hava (Eve, in transliteration), meaning 'living' or 'life-giving.' Together, therefore, Adam and Eve signified quite literally 'Soil and Life.'

The same powerful metaphor is echoed in the Latin name for the human species – Homo, derived from humus, the material of the soil. Hence, the adjective 'human' also implies 'of the soil.' Other ancient cultures evoked equally powerful associations. To the Greeks, the earth was a manifestation of Gaea, the maternal goddess who, impregnated by Uranus (god of the sky), gave birth to all the gods of the Greek pantheon.

Our civilization depends on the soil more crucially than ever, because our numbers have grown while available soil resources have diminished and deteriorated. Paradoxically, however, even as our dependence on the soil has increased, most of us have become physically and emotionally detached from it. Many of the people in the so-called 'developed' countries spend their lives in the artificial environment of a city, insulated from direct exposure to nature, and some children may now assume as a matter of course that food originates in supermarkets.

Detachment has bred ignorance, and out of ignorance has come the delusion that our civilization has risen above nature and has set itself free of its constraints. Agriculture and food security, erosion and salination, degradation of natural ecosystems, depletion and pollution of surface waters and aquifers, and decimation of biodiversity – all of these processes, which involve the soil directly or indirectly – have become abstractions to many people. The very language we use betrays disdain for that common material underfoot, often referred to as 'dirt.' Some fastidious parents prohibit their children from playing in the mud and hurry to wash their 'soiled' hands when the children nonetheless obey an innate instinct to do so. Thus soil is devalued and treated as unclean though it is the terrestrial realm's principal medium of purification, wherein wastes are decomposed and nature's productivity is continually rejuvenated.

Scientists who observe soil closely see it in effect as a seething foundry in which matter and energy are in constant flux. Radiant energy from the sun streams onto the field and cascades through the soil and the plants growing in it. Heat is exchanged, water percolates through the soil's intricate passages, plant roots extract water and transmit it to their leaves, which transpire it back to the atmosphere. Leaves absorb carbon dioxide from the air and synthesize it with soil-derived water to form the primary compounds of life. Oxygen emitted by the leaves makes the air breathable for animals, which consume and in turn fertilize plants.

Soil is thus a self-regulating bio-physio-chemical factory, processing its own materials, water, and solar energy. It also determines the fate of rainfall and snowfall reaching the ground surface – whether the water thus received will flow over the land as runoff, or seep downward to the subterranean reservoir called groundwater, which in turn maintains the steady flow of springs and streams. With its finite capacity to absorb and store moisture, and to release it gradually, the soil regulates all of these phenomena. Without the soil as a buffer, rain falling over the continents would run off entirely, producing violent floods rather than sustained river flow.

Soil naturally acts as a living filter, in which pathogens and toxins that might otherwise accumulate to foul the terrestrial environment are rendered harmless. Since time immemorial, humans and other animals have been dying of all manner of disease and have then been buried in the soil, yet no major disease is transmitted by it. The term *antibiotic* was coined by soil microbiologists who, as a consequence of their studies of soil bacteria and actinomycetes, discovered streptomycin (an important cure for tuberculosis and other infections). Ion exchange, a useful process of water purification, also was discovered by soil scientists studying the passage of solutes through beds of clay.

However unique in form and function, soil is not an isolated body. It is, rather, a central link in the larger chain of interconnected domains and processes comprising the terrestrial environment. The soil interacts both with the overlying atmosphere and the underlying strata, as well as with surface and underground bodies of water. Especially important is the interrelation between the soil and the climate. In addition to its function of regulating the cycle of water, it also regulates energy exchange and surface temperature.

When virgin land is cleared of vegetation and turned into a cultivated field, the native biomass above the ground is often burned and the organic matter within the soil tends to decompose. These processes release carbon dioxide into the atmosphere, thus contributing to the earth's greenhouse effect and to global warming. On the other hand, the opposite act of reforestation and soil enrichment with organic matter, such as can be achieved by means of conservation management, may serve to absorb carbon dioxide from the atmosphere. To an extent, the soil's capacity to store carbon can thus help to mitigate the greenhouse effect.

Thousands of years are required for nature to create life-giving soil out of sterile bedrock. In only a few decades, however, unknowing or uncaring humans can destroy that wondrous work of nature. In various circumstances, mismanaged soils may be subject to erosion (the sediments of which tend to clog streambeds, estuaries, lakes, and coastal waters), to leaching of nutrients with attendant loss of fertility and eutrophication of water bodies, to waterlogging and impaired aeration, or to an excessive accumulation of salts that may cause a once-productive soil to become entirely sterile. Such processes of soil degradation, sometimes called 'desertification,' already affect large areas of land.

We cannot manage effectively and sustainably that which we do not know and thoroughly understand. That is why the tasks of developing and disseminating sound knowledge of the soil and its complex processes have assumed growing urgency and importance. The global environmental crisis has created a compelling need for a concentrated, concise, and definitive source of information – accessible to students, scientists, practitioners, and the general public – about the soil in all its manifestations – in nature and in relation to the life of humans.

Daniel Hillel
Editor-in-Chief
May 2004

INTRODUCTION

The *Encyclopedia of Soils in the Environment* contains nearly 300 articles, written by the world's leading authorities. Pedologists, biologists, ecologists, earth scientists, hydrologists, climatologists, geographers, and representatives from many other disciplines have contributed to this work. Each of the articles separately, and all of them in sequence and combination, serve to summarize and encapsulate our present knowledge of the world's variegated soils, their natural functions, and their importance to humans.

Concise articles surveying specific aspects of soils (soil genesis, soil chemistry and mineralogy, soil physics and hydrology, and soil biology) are complemented by articles covering transdisciplinary aspects, such as the role of soils in ecology, the history of soil utilization for agricultural and engineering purposes, the development of soil science as a discipline, and the potential or actual contributions of soils to the generation, as well as to the mitigation, of pollution and of global climate change.

This comprehensive reference encompasses both the fundamental and the applied aspects of soil science, interfacing in general with the physical sciences and life sciences and more specifically with the earth sciences and environmental sciences.

The *Encyclopedia of Soils in the Environment* manifests the expanding scope of modern soil science, from its early sectarian focus on the utilitarian attributes of soils in agriculture and engineering, to a wider and much more inclusive view of the soil as a central link in the continuous chain of processes constituting the dynamic environment as a whole. Thus it both details and integrates a set of topics that have always been of vital importance to human societies and that are certain to be even more so in the future.

Daniel Hillel
Editor-in-Chief
May 2004

CONTENTS

Contents are given as follows: CHAPTER NAME Author(s) Page number

VOLUME 1

A

B

E

F

VOLUME 2

H

I

J

K

L

M

VOLUME 3

N

O

P

Q

R

S

VOLUME 4

Z

Table of Contents, Volume 2

FERTIGATION

U Kafkafi and S Kant, The Hebrew University of Jerusalem, Rehovot, Israel

Introduction

'Fertigation' is a fusion of two words: 'fertilizer' and 'irrigation.' Fertigation is the process of applying mineral fertilizers to crops along with the irrigation water. This word was first used in the USA to describe the bubbling of anhydrous ammonia into irrigation water; later, other fertilizers were applied through a sprinkler system. The widest use of fertigation is in drip irrigation. The concept of irrigation with dissolved nutrients dates back to Roman times, when city sewage was used to irrigate crops. The use of jute bags containing ammonium sulfate ($(NH_4)_2SO_4$) fertilizer at the entrance of canals and flood irrigation for banana crops has been used by growers in the Jordan Valley since the early 1930s. Mixing of micronutrients such as zinc sulfate ($ZnSO_4$) with a flood or furrow irrigation system is still in use in many developing countries, where farmers place bags of the soluble nutrients at the inlet of the irrigation water.

Fertigation in various trickle-irrigation technologies involves the injection of fertilizer solutions into irrigation systems via calibrated injection pumps. The corrosive nature of fertilizers prevents the use of fertigation when aluminum or zinc-plated metal pipes are used for irrigation. The development of plastic systems for irrigation coupled with the development of trickle-irrigation systems has enabled the refinement of fertilizer application through the irrigation lines. The fertilizer industry has adapted to field demands by introducing pure and soluble fertilizers. The time-consuming process of the dissolution of various fertilizers has led to the development of liquid fertilizer blends, according to specific recipes as ordered by growers to meet the demands of specific plants at particular growing stages and under certain climatic conditions. The use of conventional, broadcast application of fertilizers becomes ineffective with a drip-irrigation system. An adequate supply to satisfy plant demands for nutrients from a limited soil root volume can only be achieved by synchronizing the supply of water and nutrients during the various growing stages of the plants. The units used to plan fertigation are: milligrams of a nutrient consumed per day per plant, rather than kilograms of a nutrient per hectare. The daily application rate of fertigation changes during the growing season and is planned to meet the plant's daily demands according to its nutrient-uptake strategy.

Advantages of Fertigation

Fertigation has some specific advantages over broadcast and band fertilization: (1) a frequent supply of nutrients reduces fluctuation of nutrient concentration in soil; (2) there is efficient utilization and precise application of nutrients according to the nutritional requirements of the crop; (3) fertilizers are applied throughout the irrigated soil volume; (4) nutrients can be applied to the soil when soil or crop conditions would otherwise prohibit entry into the field with conventional equipment. Drip fertigation has added advantages over other methods of fertigation: (1) application of nutrients only to the wetted soil volume where roots are active reduces loss of nutrients by leaching or soil fixation and increases fertilizer-use efficiency; (2) the crop foliage remains dry, thus retarding incidence of pests or diseases and escaping foliage burn; (3) the wind has no effect and runoff is avoided.

Drip-Fertigation System

A drip-fertigation system is designed specifically for particular crop, climatic, and soil conditions.

Containers for Fertilizers

There are two main methods of delivering fertilizer solution to the irrigation lines:

1. Fertilizer dilution tanks: these are connected at the head of the irrigation line. The difference in water head between inlet and outlet ports drives some of the irrigation water through the fertilizer container and transfers it to the irrigation line;
2. External pump: the fertilizer solution is injected by positive pressure directly into the irrigation line. There are several types of pump: some use the energy of the irrigation water (the pressure and flow in the main water line is used to move a piston); others use water flow through a Venturi-type pump to suck the fertilizer solution into the irrigation line. For small areas and home gardening, accurate, electric-driven pumps are available.

The dilution tank delivers a known amount of fertilizer during the irrigation cycle until all the fertilizer is washed from the tank. In this method the concentration of the nutrients in the trickling water is varied with time, but a known total amount of nutrient per unit area is delivered. With the injection pump, it is possible to control the concentration of the nutrients in the irrigation solution.

Filtration

Filtration is a prerequisite in drip irrigation to avoid clogging of drip lines and emitters, and to maintain a predicted level of uniformity of water and fertilizer application. The main filtration system depends on the quality and source of water supply. It is used to remove gravel or sand from deep well water sources. Open surface water (ponds, rivers, or lakes) supplies may contain organic matter and algae that must be removed before entry to the lines. In a fertigation system, the second filtration step (screen filter) after the fertilizer container is necessary to remove any particulate or precipitate that comes from fertilizer mixtures. Deep water sources may contain divalent soluble iron (Fe). This iron, when it comes in contact with phosphate solution, produces a gel-like precipitate that can block the tricklers and filters. The water quality and composition must be taken into consideration at the planning stage of fertigation systems.

Distribution of Water and Nutrients in Soil

Water Flow

Two main forces, gravity and capillary, govern the movement of water in the soil during the irrigation process. During evaporation the net movement of water is in an upward vertical direction. The classic infiltration one-dimensional flow equation is:

$$\frac{\partial w}{\partial t} = \frac{\partial}{\partial z}\left[K(w)\frac{\partial H}{\partial z}\right] \qquad [1]$$

Change in water content (∂w) with time (∂t) at a position in the soil must be equal to change in specific flux in the vertical direction (z). According to the Darcy law, specific flux $= -K(w)\,\partial H/\partial z$, where K is the hydraulic conductivity, which is a function of water content (w), and $\partial H/\partial z$ is the hydraulic head gradient. Hydraulic head (H) is the sum of gravitational head (z) and matric head (h), i.e., $H = z + h$.

When water is applied from a point source by a drip system, water movement is not one-dimensional but rather is spread in vertical and radial (r) dimensions. The equation of water flow from a point source takes the form:

$$\frac{\partial w}{\partial t} = \frac{\partial}{\partial r}\left[K(w)\frac{\partial h}{\partial r}\right] + \frac{K(w)}{r} \times \frac{\partial h}{\partial r} + \frac{\partial}{\partial z}\left[K(w)\frac{\partial H}{\partial z}\right] \qquad [2]$$

In drip irrigation, water spreads from a trickler in two dimensions and creates a wetted front of different shapes (Figure 1) depending on soil type and water-discharge rate. When emitter discharge rate is higher than the soil-infiltration rate, lateral water movement dominates and a shallower penetration of water is expected when a constant amount of water is applied.

Nutrient Movement

Mass flow and diffusion govern the movement of nutrients in soil solution. Mass flow is the movement of nutrient ions with the moving water. Diffusion occurs because of concentration gradient. Two-dimensional (x and z) movement of reactive nutrient ions in soil is given by the following equation:

$$\begin{aligned}[b(c) + w]\frac{\partial c}{\partial t} = {} & \frac{\partial}{\partial x}\left[D_{\mathrm{h_{xx}}}\frac{\partial c}{\partial x} + D_{\mathrm{h_{xz}}}\frac{\partial c}{\partial z}\right] \\ & + \frac{\partial}{\partial z}\left[D_{\mathrm{h_{xz}}}\frac{\partial c}{\partial x} + D_{\mathrm{h_{zz}}}\frac{\partial c}{\partial z}\right] \\ & - \left[\frac{\partial(q_{\mathrm{x}}c)}{\partial x} + \frac{\partial(q_{\mathrm{z}}c)}{\partial z}\right] \qquad [3]\end{aligned}$$

where c is nutrient concentration in the soil solution, b is buffering capacity of the soil, q is flux of the soil solution according to the Darcy law, and D_{h} is the coefficient of diffusion. For nutrient application with a drip system, radial (r) movement of nutrients with water is added, so the nutrient flow will follow:

$$\begin{aligned}[b(c) + w]\frac{\partial c}{\partial t} = {} & \frac{\partial}{\partial r}\left[D_{\mathrm{h_{rr}}}\frac{\partial c}{\partial r} + D_{\mathrm{h_{rz}}}\frac{\partial c}{\partial z}\right] \\ & + \frac{1}{r}\left[D_{\mathrm{h_{rr}}}\frac{\partial c}{\partial r} + D_{\mathrm{h_{rz}}}\frac{\partial c}{\partial z}\right] \\ & + \frac{\partial}{\partial z}\left[D_{\mathrm{h_{zz}}}\frac{\partial c}{\partial z} + D_{\mathrm{h_{rz}}}\frac{\partial c}{\partial r}\right] \\ & - \left[\frac{\partial(q_{\mathrm{r}}c)}{\partial r} + \frac{\partial(q_{\mathrm{z}}c)}{\partial z} + \frac{q_{\mathrm{r}}c}{r}\right] \qquad [4]\end{aligned}$$

Equation [4] suggests that strongly adsorbed ions (such as phosphate; $b(c) > 1$) have lower mobility in soil than nonsorbed ions (such as nitrate or chloride). For a given nutrient concentration, $b(c)$ of fine-textured soils exceeds that of coarse-textured soils, therefore mobility of sorbed ions in clay soils is less than in sandy soils.

Salt mass movement in two different soils and two discharge rates by drip irrigation is shown in Figure 2. Both soils have the same initial salt concentration and the same salt concentration at the soil inlet. The movement of nonsorbed solute in the vertical component of the wetted zone is deeper and in the radial direction it is narrower as the soil become coarser and the emitter discharge rate becomes lower.

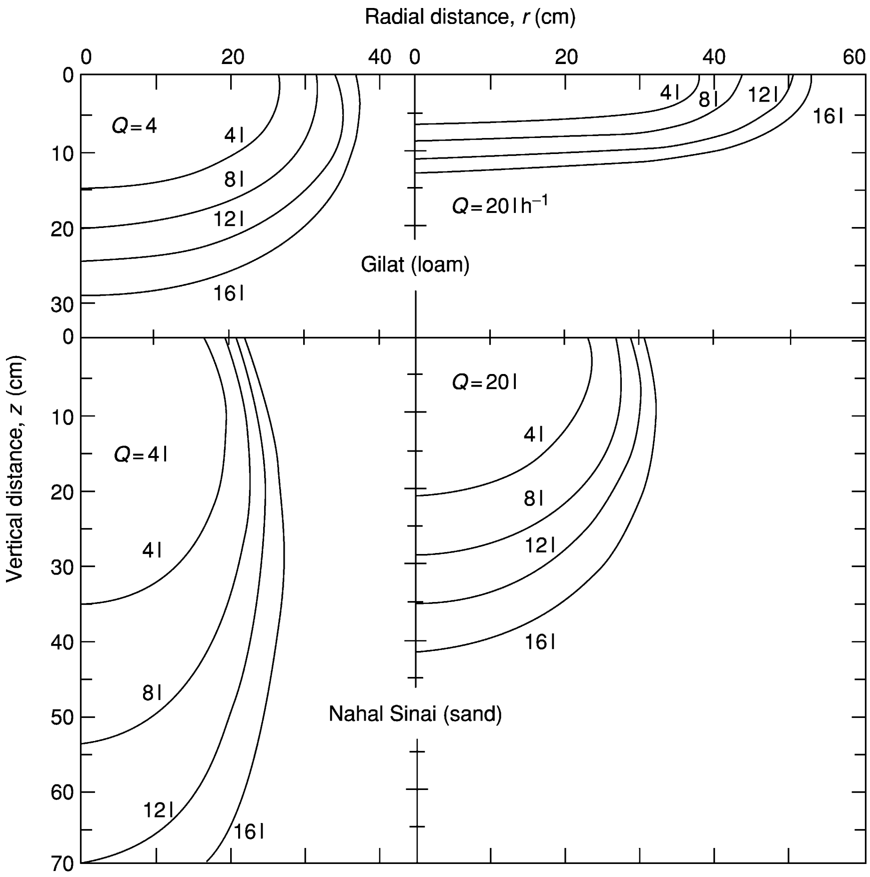


Figure 1 Water content distribution as a function of infiltration time and cumulative infiltration water (in liters) for two soils and two discharge rates (Q). (Adapted from Bresler E (1977) Trickle-drip irrigation: principles and application to soil-water management. *Advances in Agronomy* 29: 343–393, with permission.)

The concentration and movement of nonsorbed nitrate ions, added with water, follow the distribution pattern shown in Figure 2.

In repeated cycles of fertigation, there is a balance between the distance of lateral water flow and evaporation. As a result, at the border between the dry and the wet zone, a circle of soluble salts tends to accumulate, especially in hot dry areas with no intermittent rains. The salt accumulated in the wet zone periphery can reach very high levels and, in the case of a single flush of rain, this high-salt soil can be washed into the root zone and cause considerable damage. To avoid this soluble-salt accumulation on the soil surface due to evaporation, irrigation under plastic is used, especially for a permanent utilization of saline water for irrigation of vegetables and orchards. In arid climates where the evaporation rate is high, mobile nutrient ions (NO_3^-, Cl^-, Ca^{2+}, and K^+) tend to accumulate at the wet zone periphery in the soil surface. This higher concentration of soluble salts is detrimental for young seedlings, because their root systems might face high salt concentrations even with good-quality water.

Selection of Fertilizers

Most of the water-soluble and liquid fertilizers are suitable for fertigation. When choosing a fertilizer, three main considerations have to be taken into account: (1) plant type and stage of growth; (2) type of soil and irrigation system; and (3) quality of available water. Consideration has to be given to the sensitivity of a plant to the forms of N during various physiological stages. Some plants such as tomato are very sensitive to the concentration of ammonium near the root, so a nitrate-rich nutrient solution is selected for this type of plant.

At high root temperatures, ammonium might damage the roots by placing a burden on their root-soluble sugar content. In aerated soils at high temperatures, nitrate sources outweigh ammonium sources. On heavy clay soils, a zone of ponded water under the trickle point might develop where, at high soil temperatures, local anaerobic conditions can cause severe nitrate-N (NO_3^--N) losses to the atmosphere as N_2 or N_2O gases. In such conditions the plants can show N deficiency despite sufficient nitrate supply

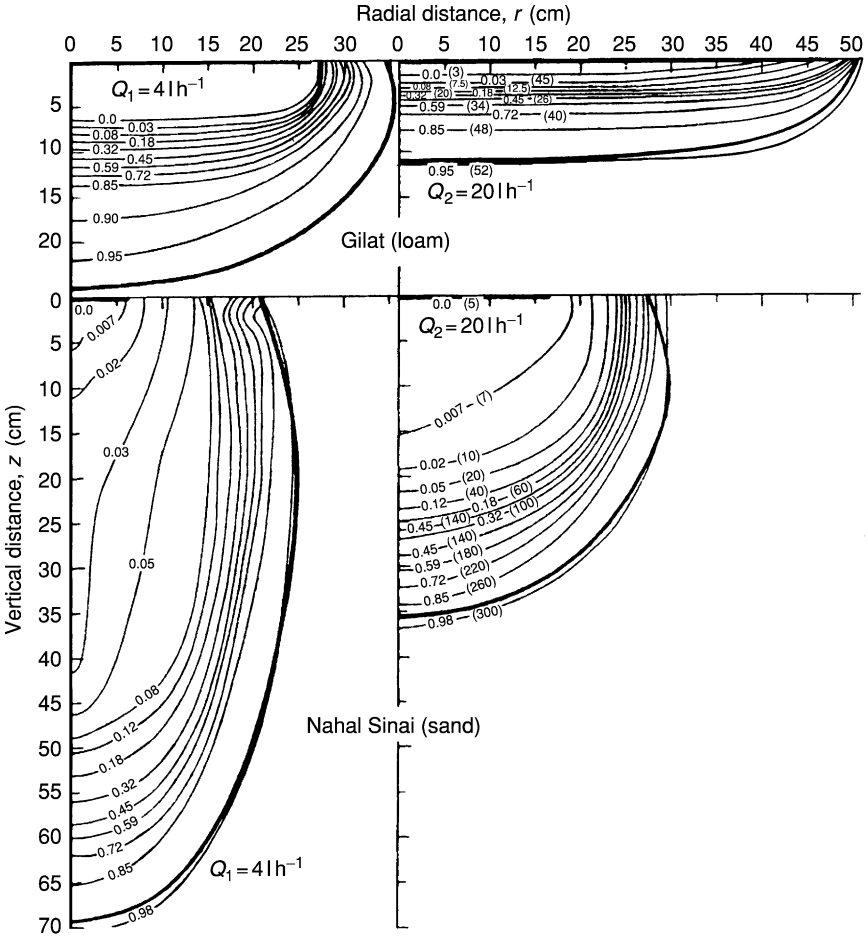


Figure 2 Distribution pattern of two-dimensional salt concentration during infiltration for two trickler discharges (Q_1 and Q_2) in loam and sandy soils. The numbers labeling the curves indicate relative salt concentration in terms of $(C - C_o)/C_n$. The numbers in parentheses are salt concentration (C) in soil solution (millimoles per liter); C_o is inlet salt concentration and C_n is initial salt concentration in soil. (Adapted from Bresler E (1977) Trickle-drip irrigation: principles and application to soil-water management. *Advances in Agronomy* 29: 343–393, with permission.)

through the irrigation line. Low concentrations of N as urea or ammonium in the irrigation solution might prevent loss and deficiency of N in plants.

Lowering the pH of the irrigation water to approximately 5.0 is often needed to keep phosphorus (P) in solution during the fertilizer injection. P application as phosphoric acid is preferable during the cold season. If micronutrients are needed, their soluble forms as chelates are less susceptible to precipitation in the irrigation lines and in the soil.

Solubility of Fertilizers

Solid fertilizers vary in their dissolution rate and in the amount that can be dissolved in a given volume of water. Solubility of fertilizers is generally reduced when two or more fertilizers are mixed together. The solubility of some commonly used fertilizers is presented in Table 1. This characteristic is crucial for the suitability of a fertilizer to be used in fertigation. Solubility of fertilizers generally increases with temperature. When present as a single salt, the main factor affecting solubility of potassium (K) and nitrate fertilizers is temperature. Whether the fertilizer dissolution is endothermic (the solution temperature is reduced during dissolution) or exothermic affects the dissolution rates and energy needed to prepare fertilizer solution. Most solid N fertilizers absorb heat when mixed with water. This results in cooling the solution and slows down the preparation of the liquid solution in the mixing tank.

The solubility of calcium nitrate ($Ca(NO_3)_2$) is higher at 30°C (Table 1); however, in practice its endothermic reaction is strong and only 50% of the amount expected to be dissolved at 30 °C is dissolved in 5 min, which results in a temperature drop to 2 °C. When the mixing tank is placed in an open field, this characteristic could reduce the expected concentration in the irrigation line, because some of the fertilizer

Table 1 Solubility, nutrient content, and soil reaction of some common fertilizers used in fertigation

					Solubility (g per 100 ml water)			
Fertilizer	*Chemical formulae*	*Grade*[a] *N–P–K*	*Other nutrients*	*Soil reaction*	*0°C*	*10°C*	*20°C*	*30°C*
Urea	$CO(NH_2)_2$	45-0-0	–	Acidic	–	78	–	–
Ammonium nitrate	NH_4NO_3	34-0-0	–	Acidic	118	152	192	242
Ammonium sulfate	$(NH_4)_2SO_4$	20-0-0	24% S	Acidic	71	73	75	78
Calcium nitrate	$Ca(NO_3)_2 \cdot 4H_2O$	15-0-0	17% Ca	Basic	102	112	121	150
Monoammonium phosphate	$NH_4H_2PO_4$	11-48-0	1.4% Ca 2.6% S	Acidic	23	29	37	46
Diammonium phosphate	$(NH_4)_2HPO_4$	21-53-0	–	Acidic	43	–	–	–
Phosphoric acid	H_3PO_4	0-52-0	–	Very acidic	548	–	–	–
Muriate of potash	KCl	0-0-62	–	Neutral	28	31	34	37
Potassium nitrate	KNO_3	13-0-44	–	Basic	13	21	31	45
Potassium sulfate	K_2SO_4	0-0-53	8% S	Neutral	7	9	11	13

[a] Concentrations of N, P_2O_5, and K_2O, expressed as a percentage of weight of the fertilizer.

will remain as a solid in the tank. The diverse solubility characteristics of fertilizers and the problems they cause to the end user were the trigger to the establishment of dilution plants that supply to the farmers a nutrient cocktail to order, to meet specific plant needs. Another approach to the solubility problem of a mixed nutrient solution has been adopted in the modern glasshouse industry, where different tanks of separate nutrient sources and acid and bases are used. With the help of a computer, it is possible to program the amounts of each nutrient element to be pumped into the irrigation line in small doses. This technology is too expensive to operate in large, open-field and plantation crops, where the Venturi or proportional pumps are used to inject the dose of fertilizer solution from a storage tank into the irrigation line.

Urea, ammonium nitrate, calcium nitrate, potassium nitrate, and ammonium phosphate are easily soluble in water and are used extensively in the preparation of single-nutrient or multinutrient fertilizer solutions. Monoammonium phosphate, phosphoric acid, and urea phosphate are water-soluble, but they may have precipitation problems when injected at high rates into hard water (containing large amounts of calcium and magnesium carbonates). All K fertilizers are water-soluble, but they vary in their rates of dissolution and sensitivity to the solution temperature. A common field problem in tanks of mixed fertilizer solution is precipitation during cold nights.

Compatibility of Fertilizers

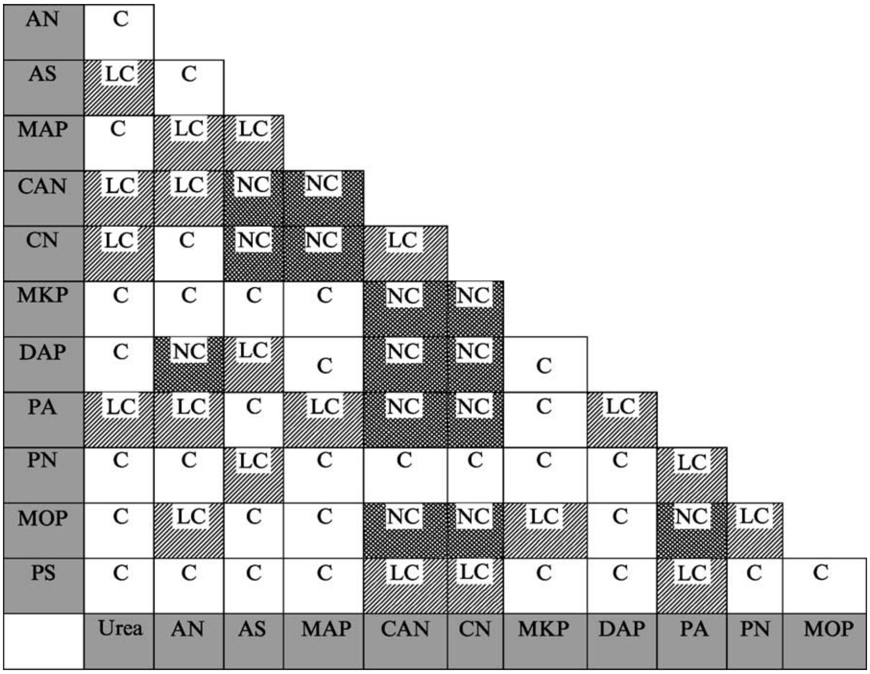
Interaction between fertilizers can affect their availability and uptake by plants. Mixing of two fertilizers can sometimes result in precipitation. Injection of Ca salts with phosphate or sulfate may increase the chances of precipitate formation as calcium phosphate or calcium sulfate ($CaSO_4$). Fertilizer compatibility is presented in Figure 3. The pH of the irrigation solution should be within the range 5.5–7.0. A pH that is too high will reduce the availability of P, zinc, and iron and may result in precipitation of Ca and Mg phosphates or carbonates in the irrigation lines. A pH value that is too low is detrimental to roots and may increase Al and Mn concentrations in the soil solution. Nitric (HNO_3) or phosphoric (H_3PO_4) acids are used to reduce the pH level in fertigation. The advantage of using these acids is that, beside dissolution of basic precipitates in the line, they also supply the plants with the essential nutrients nitrate and phosphate, replacing other fertilizers containing N and P. In saline water and calcareous clay soils, nitric acid minimizes the increased salinity in the root zone and supplies only nitrate, which counterbalances excess chloride.

Precipitation in the Irrigation Lines

Precipitation of insoluble dicalcium phosphate and dimagnesium phosphate and calcium carbonate can develop with high-pH water. Iron phosphate precipitate can form in drip lines with low-pH water. Water containing high levels of Mg ions may cause an ammonium magnesium phosphate precipitation in the fertilizer tank. Avoiding the use of ammonium fertilizer in such conditions can prevent the risk of blocking the emitters. Using potassium sulfate (K_2SO_4) or ammonium sulfate ($(NH_4)_2SO_4$) with water containing high concentrations of Ca can result in precipitation as $CaSO_4$ (gypsum precipitates), which can clog the drip lines.

Scheduling Fertigation

The nutrient elements are taken up by plants depending upon the plants' requirements at a specific developmental stage. Injecting fertilizer into a



	Urea	AN	AS	MAP	CAN	CN	MKP	DAP	PA	PN	MOP
AN	C										
AS	LC	C									
MAP	C	LC	LC								
CAN	LC	LC	NC	NC							
CN	LC	C	NC	NC	LC						
MKP	C	C	C	C	NC	NC					
DAP	C	NC	LC	C	NC	NC	C				
PA	LC	LC	C	LC	NC	NC	C	LC			
PN	C	C	LC	C	C	C	C	C	LC		
MOP	C	LC	C	C	NC	NC	LC	C	NC	LC	
PS	C	C	C	C	LC	LC	C	C	LC	C	C

Figure 3 Compatibility of common fertilizers. C, compatible; LC, limited compatibility; NC, not compatible; AN, ammonium nitrate; AS, ammonium sulfate; MAP, monoammonium phosphate; CAN, calcium ammonium nitrate; CN, calcium nitrate; MKP, monopotassium phosphate; DAP, diammonium phosphate; PA, phosphoric acid; PN, potassium nitrate; MOP, muriate of potash (KCl).

drip-irrigation system offers the benefits of nutrient application to the crop in optimum amounts and at times when it is most likely to be taken up by the plant. Fertigation provides optimum concentrations and quantities of nutrients directly into the root zone. Hence, accurate planning of when and how much fertilizer to apply is critical for fertigation management. Nutrient requirements by plants change with their physiological growth stages. The climatic factors, timing of fertilizer application, and irrigation timings affect crop growth and consequently affect the quantity of nutrients required. A fertigation schedule depends upon soil characteristics and fertilizer requirements of the growing plants. Nutrient (N, P, and K) uptake by some field and vegetable crops with respect to growth is given in Tables 2–4. Soil type, system design, and length of growing season govern the frequency of fertilizer application. The frequency of injecting fertilizers varies with soil type and specific plant requirements. In plants grown on sand dunes, several irrigations per week might be needed, while on heavy clay soils only one or two irrigations per week are needed. The smaller the root volume, the higher is the fertigation frequency.

Behavior of Nutrients in Soil

Soil chemical properties are important in planning fertigation. pH has a great effect on the availability of residual nutrients in soil as well as on those added via fertigation. The balance between cation and anion uptake by the plant affects the pH in the rhizosphere. When a plant takes up more nutrient cations, such as when NH_4^+ is the main N source, protons are exuded by the roots and acidify the soil surrounding the roots. If the anion uptake is predominant, as when NO_3^- is the main source of N, the roots exude OH^- or HCO_3^-, resulting in a pH rise in the rhizosphere. Nitrate and ammonium are the main forms of N available for plant uptake. Ammonium nutrition leads to excretion of 0.9 mol of H^+ by roots with 1 mol of ammonium assimilation; while, where there is nitrate reduction in the leaves, 0.1–0.3 mol of OH^- is released from the root per mole of NO_3^- taken up. The rhizosphere pH can change with the form and concentration of the N fertilizer, but the extent of the pH change around the root depends on the buffer capacity of the soil.

Cation exchange capacity (CEC) of the soil is an important consideration in determining the amount and frequency of cations that need to be added during fertigation. In most agricultural soils and irrigation water, Ca and Mg are present in larger quantities than those taken up by any crop, and their supply is usually satisfied with water mass flow. K^+ is the main cation that must be supplied with the irrigation water. In order to maintain an acceptable concentration of K in the soil solution, a soil with low CEC must

Table 2 Nitrogen uptake by various field and vegetable crops with respect to relative time of growth

Crop	*Relative time of growth (%)*					*Total uptake*	*Plants ha^{-1}*	*Projected yield*
	0–20	*20–40*	*40–60*	*60–80*	*80–100*	*(g plant^{-1})*		*(t ha^{-1})*
Cotton								
g plant^{-1}	0.20	1.80	3.80	2.20	1.60	9.60	25 000	1.3[b]
mg N plant^{-1}day^{-1}[a]	6	58	123	71	52	62		
Maize								
g plant^{-1}	0.25	1.58	1.00	0.83	0.50	4.17	60 000	8
mg N plant^{-1}day^{-1}[a]	11	70	44	37	22	37		
Tomato								
g plant^{-1}	0.50	0.75	2.50	4.25	3.25	11.25	20 000	100
mg N plant^{-1}day^{-1}[a]	19	28	91	156	119	83		
Sweet pepper								
g plant^{-1}	0.40	1.80	1.10	0.70	0.60	4.60	50 000	55
mg N plant^{-1}day^{-1}[a]	20	90	55	35	30	46		
Potato								
g plant^{-1}	0.08	1.00	1.08	0.50	0.17	2.83	60 000	50
mg N plant^{-1}day^{-1}[a]	4	50	54	25	9	28		
Muskmelon								
g plant^{-1}	0.20	0.60	1.60	2.80	0.80	6.00	25 000	50
mg N plant^{-1}day^{-1}[a]	10	30	80	140	40	60		
Watermelon								
g plant^{-1}	0.83	1.67	3.33	6.67	2.50	15.00	12 000	75
mg N plant^{-1}day^{-1}[a]	41	84	166	333	125	150		
Cabbage								
g plant^{-1}	0.10	0.20	0.80	1.90	0.60	3.60	50 000	29
mg N plant^{-1}day^{-1}[a]	8	16	63	150	47	56		
Cauliflower								
g plant^{-1}	0.10	0.20	0.50	2.00	1.40	4.20	50 000	9
mg N plant^{-1}day^{-1}[a]	8	16	40	157	110	66		
Eggplant								
g plant^{-1}	0.50	3.25	2.00	2.50	1.50	9.75	20 000	40
mg N plant^{-1}day^{-1}[a]	14	89	55	69	41	54		

[a] Daily amounts of nitrogen (mg N plant^{-1} day^{-1}) added with drip fertigation; this amount includes 10% additional nitrogen for root consumption.

[b] Seed cotton (lint) yield.

get a fresh supply of K more frequently than in soil with high CEC, which can hold higher quantities of K. Fertigation is most practical in sandy soils and soils of dry and arid regions that have a low CEC, because these soils need frequent irrigation and quick nutrient replenishment. Old farming practices considered sand dunes as nonagricultural soils, but the introduction of fertigation has turned desert sand dunes into productive agricultural soils.

Nitrate-N is highly mobile and is more likely to be lost through surface runoff and leaching. In trickle irrigation, ponding under the trickler creates an oxygen-devoid space in which denitrification of applied nitrate is observed during the irrigation cycle. The rate of water discharge from an emitter is chosen to meet the rate of water entry to the soil from a point source. Volatilization of ammonia during urea application might also occur if the irrigation solution is basic. Acidifying the irrigation water prevents ammonia losses.

Added phosphate is rapidly adsorbed and precipitated in the soil, leading to a rapid decline in water-soluble phosphate concentration in the soil solution. Its movement is retarded owing to retention by soil oxides and clay minerals. P application with drip irrigation is more efficient than with sprinkler irrigation or broadcasting. P in fertigation is supplied directly into the zone where active roots are present, allowing its immediate uptake by active roots before drying and irreversible fixation occur in the soil.

Root Growth

To achieve optimum plant growth, the root zone must be well supplied with water, nutrients, and oxygen. Water potential kept by frequent irrigation at a consistently low water tension, especially in clay soil, can result in a suboptimal supply of oxygen in the root zone. Roots respond within minutes to a reduction in oxygen supply by cessation of root extension. The elongation zone of a cotton root dies after only 30 min without oxygen. In drip irrigation the soil is saturated only in the vicinity of the trickler (Figure 1), while most of the soil volume is wetted by an

Table 3 Phosphorus uptake by various field and vegetable crops with respect to relative time of growth

Crop	*Relative time of growth (%)*					*Total uptake (g plant^{-1})*	*Plants ha^{-1}*	*Projected yield (t ha^{-1})*
	0–20	*20–40*	*40–60*	*60–80*	*80–100*			
Cotton								
g plant^{-1}	0.17	0.24	0.80	0.44	0.17	1.80	25 000	1.3[b]
mg P_2O_5 plant^{-1} day^{-1}[a]	5.2	7.7	25.8	14.2	5.2	11.6		
Maize								
g plant^{-1}	0.07	0.30	0.28	0.25	0.10	1.00	60 000	8
mg P_2O_5 plant^{-1} day^{-1}[a]	2.9	13.2	12.1	11.0	4.4	8.8		
Tomato								
g plant^{-1}	0.03	0.05	0.17	0.45	0.25	0.95	20 000	100
mg P_2O_5 plant^{-1} day^{-1}[a]	1.1	1.8	6.2	16.5	9.0	7.0		
Sweet pepper								
g plant^{-1}	0.03	0.10	0.20	0.08	0.04	0.45	50 000	55
mg P_2O_5 plant^{-1} day^{-1}[a]	1.5	5.0	10.0	4.0	2.0	4.5		
Potato								
g plant^{-1}	0.01	0.05	0.10	0.14	0.09	0.39	60 000	50
mg P_2O_5 plant^{-1} day^{-1}[a]	0.5	2.5	5.0	7.0	4.5	3.9		
Muskmelon								
g plant^{-1}	0.02	0.08	0.20	0.32	0.20	0.82	25 000	50
mg P_2O_5 plant^{-1} day^{-1}[a]	1.1	4.0	10.0	16.0	10.0	8.2		
Eggplant								
g plant^{-1}	0.03	0.12	0.18	0.42	0.35	1.10	20 000	40
mg P_2O_5 plant^{-1} day^{-1}[a]	0.8	3.3	5.0	11.5	9.6	6.0		

[a] Daily amounts of phosphorus (mg P_2O_5 plant^{-1} day^{-1}) to be added with drip fertigation; this amount includes 10% additional phosphorus for root consumption.

[b] Seed cotton (lint) yield.

Table 4 Potassium uptake by various field and vegetable crops with respect to relative time of growth

Crop	*Relative time of growth (%)*					*Total uptake (g plant^{-1})*	*Plants ha^{-1}*	*Projected yield (t ha^{-1})*
	0–20	*20–40*	*40–60*	*60–80*	*80–100*			
Cotton								
g plant^{-1}	0.60	2.00	3.60	0.60	0.20	7.00	25 000	1.3[b]
mg K_2O plant^{-1} day^{-1}[a]	20	65	117	20	7	45		
Maize								
g plant^{-1}	0.25	1.83	1.00	0.33	0.08	3.50	60 000	8
mg K_2O plant^{-1} day^{-1}[a]	11	80	44	14	4	31		
Sugarcane								
g plant^{-1}	0.50	0.60	0.70	1.80	0.60	4.20	50 000	140
mg K_2O plant^{-1} day^{-1}[a]	11	13	15	40	13	19		
Tomato								
g plant^{-1}	0.70	0.80	3.50	7.00	4.50	16.50	20 000	100
mg K_2O plant^{-1} day^{-1}[a]	25	30	128	256	165	121		
Sweet pepper								
g plant^{-1}	0.50	2.00	1.40	1.40	0.40	5.70	50 000	55
mg K_2O plant^{-1} day^{-1}[a]	25	100	70	70	20	57		
Potato								
g plant^{-1}	0.20	0.80	1.80	1.50	0.40	4.70	60 000	50
mg K_2O plant^{-1} day^{-1}[a]	10	40	90	75	20	47		
Muskmelon								
g plant^{-1}	0.40	1.20	4.00	4.40	2.00	12.00	25 000	50
mg K_2O plant^{-1} day^{-1}[a]	20	60	190	220	100	120		
Eggplant								
g plant^{-1}	0.75	5.00	3.00	1.75	1.00	11.50	20 000	40
mg K_2O plant^{-1} day^{-1}[a]	21	138	82	48	28	64		

[a] Daily amounts of potassium (mg K_2O plant^{-1} day^{-1}) to be added with drip fertigation; this amount includes 10% additional potassium for root consumption.

[b] Seed cotton (lint) yield.

unsaturated flow. This situation arises when tricklers of high discharge rates (more than $1\,\mathrm{l\,h^{-1}}$) are used in clay soils. Oxygen is excluded from the saturation zone when there is a continuous supply of water at high flux. A slow flow rate can maintain the moisture and oxygen regime in the wet soil volume at optimum conditions.

The nitrate-to-ammonium ratio has a significant influence on the development of the root system. A higher concentration of ammonium has deleterious effects on root growth specifically when root temperature is high. At a high root temperature, additional sugar is required simultaneously in large amounts for respiration and ammonium metabolism in the root. The supply of sugar from the leaves lags behind the respiration and metabolism consumption, resulting in momentary excess of ammonia in the root that causes root death. Under the same high-temperature conditions, nitrate reduction occurs mainly in leaves, where the sugar supply is near the site of N metabolism. At high soil temperatures, especially in soilless cultures, nitrate-N is a safe form of N during fertigation. At low root temperatures, a high level of NO_3^- accumulates in the roots and low amounts are transported to the shoot, resulting in N deficiency. Therefore, the proper concentration and form of N applied in fertigation should be chosen differently for winter and summer seasons for intensively grown crops.

Subsurface Fertigation

Installation and disassembling of laterals in drip fertigation every season require extra labor and energy. Damage to drip lines by movement of heavy machinery and animals restricts the use of surface drip-irrigation, although placement of trickle lines at appropriate soil depths can solve such problems. Recent improvements in liquid, soluble, and compatible fertilizers, filtration devices, and reduced clogging problems have opened the way for development of subsurface fertigation. Special slow-release chemicals imbedded in the plastic of the filters prevent root entry into the trickler. In addition to cost-effectiveness and energy saving, subsurface drip-fertigation has added agronomic advantages over surface drip-fertigation: (1) placement of nutrients in the region where root activity is maximal and the daily and seasonal temperature fluctuations are minimal. Distribution of nutrients in subsurface applications is more uniform and spherical around emitters than surface fertigation, where movement is hemispheric below the point source and evaporation causes accumulation of soluble salts at the wet zone periphery; (2) roots can explore a greater soil volume for water and nutrients, since roots grow deeper in a subsurface system; (3) the top 4–5 cm of soil remains dry in subsurface fertigation, thereby reducing evaporation losses and weed germination.

Fertigation gives the plant grower a tool to control the daily supply of water and nutrients in tandem with plant growth rate and nutrient demand. When properly employed, plant potential growth can be achieved that is only limited by radiation and temperature. To reach maximum plant performance, the accurate design of fertigation system, efficient utilization of equipment, combined with the timely application and selection of compatible fertilizers, are essential criteria. More detailed research on surface and subsurface fertigation is needed for different commercial crops. Fertigation in protected agriculture and potted plants must take into consideration the root zone temperature when choosing N fertilizer. Further study on the availability and movement of dissolved nutrients in different soil types to meet the nutrient requirements of important crops as a function of time is needed for open-field-grown crops.

See also: **Fertilizers and Fertilization**; **Irrigation:** Methods; **Nutrient Availability**; **Nutrient Management**

Further Reading

Bar Yosef B (1999) Advances in fertigation. *Advances in Agronomy* 65: 1–77.

Bresler E (1977) Trickle-drip irrigation: principles and application to soil-water management. *Advances in Agronomy* 29: 343–393.

Burt C, O'Connor K, and Ruehr T (1998) *Fertigation.* CA: Irrigation Training and Research Center.

Dasberg S and Bresler E (1985) *Drip Irrigation Manual.* Bet Dagan, Israel: International Irrigation Information Center.

Hanan JJ (1998) *Greenhouses.* New York: CRC Press.

Haynes RJ (1985) Principles of fertilizer use for trickle irrigated crops. *Fertilizer Research* 6: 235–255.

Hochmuth GJ (1995) *Fertigation of Vegetable Crops in Florida, USA*, pp. 199–214 Dahalia Greidinger International Symposium on Fertigation, 26 Mar–1 Apr, 1995. Haifa, Israel: Technion-Israel Institute of Technology.

Kafkafi U (1994) Combined irrigation and fertilization (fertigation) in arid zones and protected agriculture. *Israel Journal of Plant Science* 42: 301–320.

Mmolawa K and Or D (2000) Root zone solute dynamics under drip irrigation: a review. *Plant and Soil* 222: 163–190.

Scaife A and Bar-Yosef B (1995) *Fertilizing for High-Yield and -Quality Vegetables.* IPI Bulletin 13. Basel, Switzerland: International Potash Institute.

Sneh M (1995) *The History of Fertigation in Israel*, pp. 1–10. Dahalia Greidinger International Symposium on Fertigation, 26 Mar–1 Apr, 1995. Haifa, Israel: Technion-Israel Institute of Technology.

FERTILITY

J L Havlin, North Carolina State University, Raleigh, NC, USA

Introduction

Fertility of the soil is characterized by physical, chemical, and biological processes that control plant nutrient availability. Knowledge of these processes and interactions is critical to optimizing nutrient supply and plant productivity. Plants absorb nutrients dissolved in the soil solution. Nutrients in plant residue left in the field after harvest are ultimately returned to the soil and are subject to the same reactions as fertilizer nutrients. Although this cycle varies among nutrients, understanding nutrient dynamics in the soil–plant–atmosphere system is essential to successful nutrient management.

The Importance of Soil Fertility

World population will reach 9.3 billion people by 2050. Since additional arable land is only 15–20% more than current cultivated land, continued increases in productivity are essential to meet future food and fiber needs. While crop productivity has increased greatly since 1950, further advances in agricultural technologies must occur to enhance crop yield. Principal factors contributing to higher yields include development of improved cultivars, nutrient (soil fertility) and pest management, soil and water conservation, and cultural practices. Development and use of fertilizer and pesticides are directly related to increased crop productivity. Without effective and efficient management of soil fertility, 30–40% more land area would be needed today to produce the same yields. Therefore, growers must take advantage of technologies that increase productivity.

Maximum crop production depends on the growing-season environment and the ability of the producer to minimize all yield-limiting factors through appropriate management technologies. A well-established principle called the Law of the Minimum requires that the most yield-limiting factor be corrected first, the second most limiting factor next, and so on. For example, the correct variety planted at the optimum time and population may not result in the highest yield if plant-available water or nutrient availability were the most limiting factors. Thus, minimizing water as a yield-limiting factor must be accomplished before the crop can fully respond to management of any other factor.

Essential Plant Nutrients

An essential plant nutrient is involved in metabolic functions such that the plant cannot complete its life cycle without the nutrient. Seventeen nutrients are essential to plant growth (Table 1). Photosynthesis converts CO_2 and H_2O into carbohydrates that form proteins and other compounds. Macronutrients

Table 1 Essential plant nutrients and their relative and average plant concentrations

	Plant concentration[a]		
Element or nutrient	*Relative*	*Average*	*Classification*
Hydrogen H	60 000 000	6.0%	
Oxygen O	30 000 000	45.0%	
Carbon C	30 000 000	45.0%	
Nitrogen N	1 000 000	1.5%	
Potassium K	400 000	1.0%	Primary Macronutrients
Phosphorus P	30 000	0.2%	
Calcium Ca	200 000	0.5%	
Magnesium Mg	100 000	0.2%	
Sulfur S	30 000	0.1%	
Chloride Cl	3000	100 ppm (0.01%)	
Iron Fe	2000	100 ppm	
Boron B	2000	20 ppm	
Manganese Mn	1000	50 ppm	Micronutrients
Zinc Zn	300	20 ppm	
Copper Cu	100	6 ppm	
Molybdenum Mo	1	0.1 ppm	
Nickel Ni	1	0.1 ppm	

[a] Concentration expressed on a dry-matter-weight basis.

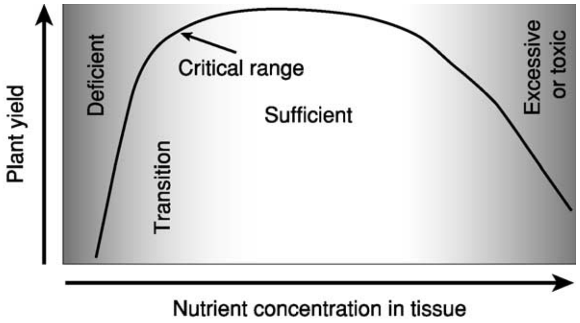


Figure 1 Relationship between plant nutrient concentration and yield. Nutrient-deficient plants exhibit visual deficiency symptoms (abnormal leaf color or other growth characteristic) and respond in yield to addition of the deficient nutrient. Plant nutrient concentration below which plant growth or yield is increased with nutrient addition is the critical level or range. Sufficiency or luxury consumption is the concentration range where added nutrient does not increase yield but increases nutrient concentration. Excessive or toxic nutrient concentrations reduce plant growth and yield.

and micronutrients are classified by their relative abundance in plants (Table 1). Sodium (Na), cobalt (Co), vanadium (Va), and silicon (Si) are beneficial in some specific plants.

When plant nutrient concentration is deficient enough to reduce plant growth severely, distinct visual deficiency symptoms appear (Figure 1). Extreme deficiencies can result in plant death. Visual symptoms may not appear when a nutrient is marginally deficient, but plant yield may still be reduced. Plant nutrient concentration above which plant growth or yield is not increased is the critical level or range. Critical nutrient ranges vary among plants but always occur in the transition between nutrient deficiency and sufficiency (Figure 1). Sufficiency or luxury consumption is the concentration range where added nutrient does not increase yield but increases nutrient concentration. Nutrient concentration is considered excessive or toxic when plant growth and yield are reduced. Plants also absorb many nonessential elements in the soil solution. For example, high Al^{3+} in plants occurs in acid soils with relatively large amounts of soluble Al. In severely acid soils Al toxicity can reduce plant yield.

Basic Soil Properties and Processes

Plant nutrient availability is a dynamic interaction of many soil processes (Figure 2). As plants absorb nutrients from soil solution, reactions occur to resupply the solution. Specific nutrients are influenced more by some reactions than by others. For example, biological processes are more important to N and S availability, and surface exchange reactions are important for K, Ca, and Mg.

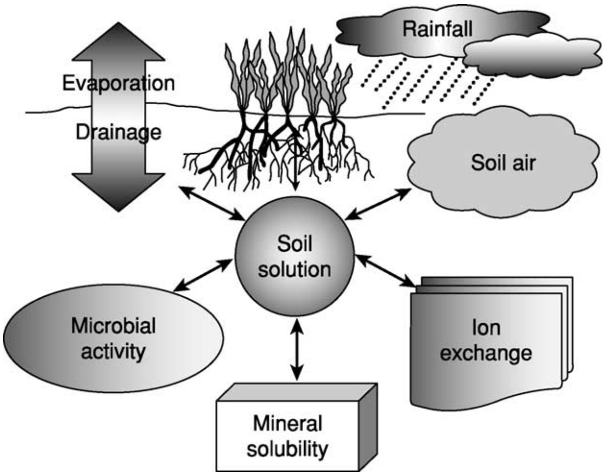


Figure 2 General diagram of nutrient cycling in soil. See text for description.

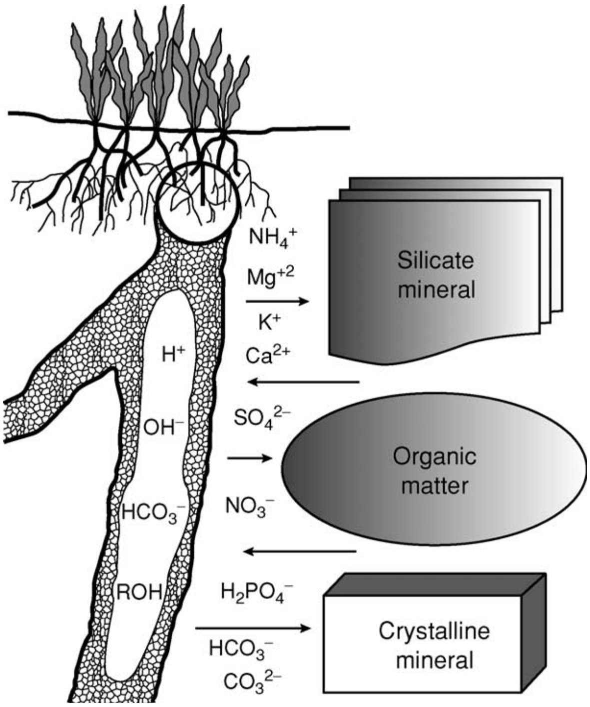


Figure 3 Ion exchange between nutrient sources (minerals and OM) and plant roots. As roots absorb nutrients from soil solution, nutrients adsorbed to the cation or anion exchange capacity are desorbed, nutrient-bearing minerals dissolve, and/or OM mineralization resupplies nutrients to soil solution.

Cation and Anion Exchange

Exchange of cations and anions on surfaces of clay minerals, inorganic compounds, organic matter (OM), and roots is one of the most important soil chemical properties influencing nutrient availability (Figure 3). Adsorbed ions are reversibly exchanged with other ions in solution. Cation exchange capacity (CEC) represents the quantity of negative charge

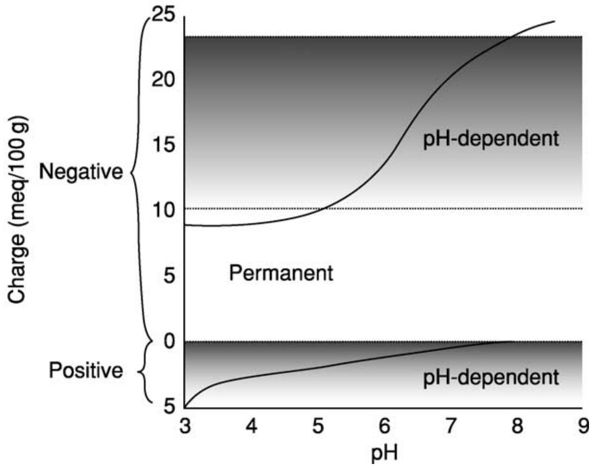


Figure 4 Influence of soil solution pH on CEC and AEC. Permanent charge is due to isomorphic substitution and pH-dependent charge is due to broken edges.

Table 2 Typical range in cation exchange capacities (CEC) for different-textured soils

Soil texture	*CEC (cmol kg^{-1})*
Sands (low OM)	3–5
Sands (high OM)	10–20
Loams	10–15
Silt loams	15–25
Clay and clay loams	20–50
Organic soils	50–100

available to attract cations. Anion exchange capacity (AEC) represents the positive charge available to attract anions in solution. In most soils CEC > AEC.

CEC generated through isomorphic substitution is permanent charge unaffected by solution pH (Figure 4). Both CEC and AEC also occur on mineral edges and OM surfaces, and vary with pH (Figure 4). Soils with greater clay and OM contents have a higher CEC (and AEC) than sandy, low-OM soils (Table 2).

CEC greatly influences nutrient availability, as most adsorbed cations are plant nutrients, except for Al^{3+}. In acid soils the primary cations are Al^{3+}, H^+, Ca^{2+}, Mg^{2+}, and K^+, whereas when soil pH ≥ 6.5 major exchangeable cations are Ca^{2+}, Mg^{2+}, K^+, and Na^+. Cations are adsorbed with different strengths, which influences the relative ease of desorption. The lyotropic series, or relative strength of adsorption, is $Al^{3+} > H^+ > Ca^{2+} > Mg^{2+} > K^+ = NH_4^+ > Na^+$. Cations with greater charge are more strongly adsorbed. For cations with similar charge, adsorption strength is determined by the size of the hydrated cation; greater hydrated cation radii reduce adsorption because larger hydrated cations cannot get as close to the exchange site as smaller cations. For anions, the relative strength of adsorption is $H_2PO_4^{2-} > SO_4^{2-} \gg NO_3^- = Cl^-$. In most soils $H_2PO_4^{2-}$ is the primary anion adsorbed, although in severely acid soils SO_4^{2-} adsorption represents a major source of plant available S.

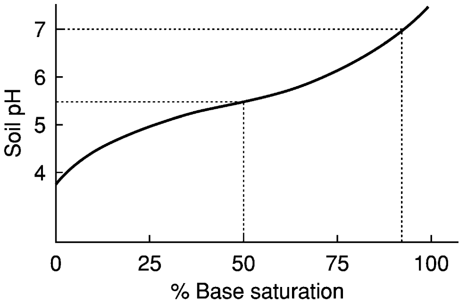


Figure 5 General relationship between soil pH and base saturation.

Base Saturation

Base saturation (BS) represents the percentage of CEC occupied by bases (Ca^{2+}, Mg^{2+}, K^+, and Na^+). The %BS increases with increasing soil pH (Figure 5). The availability of Ca^{2+}, Mg^{2+}, and K^+ increases with increasing %BS. For example, an 80% BS soil provides cations to plants more easily than 40% BS soil. At pH 5.5 most soils have 45–55% BS, whereas at pH 7 BS is ≥90%.

Buffer Capacity

Plant nutrient availability depends on nutrient concentration in solution and on the buffer capacity (BC). For example, as roots absorb K^+, K^+ on the CEC is desorbed to resupply solution K^+. With P, adsorbed $H_2PO_4^{2-}$ will desorb, while P minerals may also dissolve to resupply solution $H_2PO_4^{2-}$. The ratio of adsorbed nutrient quantity (Q) to solution nutrient intensity (I) describes soil BC:

$$\mathrm{BC} = \frac{\Delta Q}{\Delta I}$$

In Figure 6, $\mathrm{BC}_{\mathrm{soil\ A}} > \mathrm{BC}_{\mathrm{soil\ B}}$. Thus, increasing adsorbed nutrients increases solution concentration in soil B much more than soil A. As plant uptake decreases nutrients in solution (ΔI), soil A more readily buffers solution concentration than soil B. Since $\mathrm{BC}_{\mathrm{soil\ A}} > \mathrm{BC}_{\mathrm{soil\ B}}$, soil A likely contains greater clay and/or OM, resulting in greater CEC. In low BC soils where exchangeable K^+ might be too low to provide sufficient plant-available K, K must be added to increase exchangeable K to prevent K deficiency. Thus, BC strongly influences nutrient availability.

Soil Organic Matter

Soil OM represents organic materials in various stages of decay. The most important fraction to soil fertility is humus that is relatively resistant to microbial degradation, but is essential for maintaining optimum soil physical conditions important for plant growth (Figure 7). Fresh crop residues undergo rapid microbial decomposition. The heterotrophic biomass represents soil microorganisms and fauna responsible for mineralization and immobilization processes that influence availability of N, P, S, and other nutrients. Residue decomposition and humus formation processes depend on climate, soil type, and soil and crop management practices. Soil OM can have substantial effects on soil processes and properties that influence nutrient availability (Table 3). For example, increasing soil OM increases soil aggregation and decreases bulk density, which improves water infiltration, air exchange, root proliferation, and crop productivity. Soil OM levels depend on soil and crop management. If management is changed, a new OM level is attained that may be lower or higher than the previous level. Management systems should be adopted that sustain or increase soil OM and productivity. Proper nutrient management will produce high yields, which will increase the quantity of residues and soil OM.

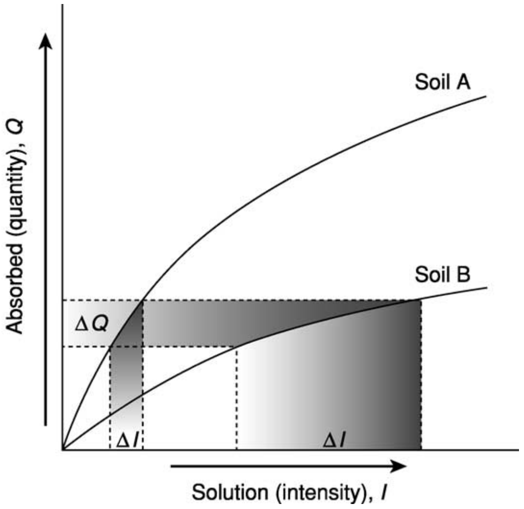


Figure 6 The relationship of adsorbed and solution nutrients as influenced by soil type. As nutrient concentration in solution (ΔI) decreases, soil A more readily buffers the solution concentration than soil B. Since $BC_{\text{soil A}} > BC_{\text{soil B}}$, soil A likely contains greater clay and/or OM content, resulting in greater CEC.

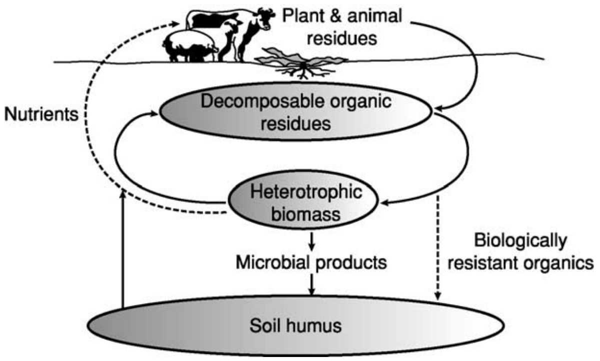


Figure 7 General degradation process of crop residue or other organic material added to soil.

Soil Acidity and Alkalinity

Acid soils usually occur in regions where annual precipitation >600–800 mm. Optimum soil pH varies between 4.5 and 7.5 depending on the crop. Neutralizing soil acidity by liming to ≥pH 5.6 will reduce exchangeable Al^{3+} to <10% of the CEC and will usually prevent Al toxicity. Decreasing soil pH under cropping systems is attributed to: (1) use of acid-forming fertilizers and manures; (2) crop removal and/or leaching of cations that decreases %BS; and (3) decomposition of organic residues. Soil acidification is enhanced with increasing rainfall since rain has a pH of 5.7 or less, depending on pollutants such as SO_2, NO_2, and others. Soil pH can strongly influence nutrient availability depending on the specific nutrient and the magnitude of the pH change. These influences will be discussed below under each specific nutrient.

Calcareous soils contain solid $CaCO_3$, exhibit soil pH ≥7.2, and occur in regions of <500 mm annual precipitation. Increasing rainfall increases depth to $CaCO_3$ where no $CaCO_3$ exists in the rooting zone at >800–1000 mm annual rainfall. If $CaCO_3$ exists, all of it must be dissolved or neutralized before soil pH could decrease. In most situations, reducing soil pH by neutralizing $CaCO_3$ is not practical. High soil

Table 3 Soil organic matter (OM) influences on soil properties and processes

Property	*Remarks*	*Effect on soil*
Soil color	Increasing OM darkens color	Darker color causes soils to warm up faster
H_2O-holding capacity	OM adsorbs ~20 times weight in H_2O	Enhances plant-available water
Soil aggregation	Binds clay particles into aggregates	Aggregation enhances permeability to H_2O/air
Chelation	Forms complexes with cations	Enhances micronutrient availability
Ion exchange/buffering	OM exhibits both CEC and AEC	Higher buffer capacity increases nutrient supply
Mineralization	OM degradation releases nutrients	Important nutrient source
Water solubility	OM is relatively insoluble in H_2O	OM does not extensively leach
Organic complexation	OM complexes organic compounds	Influences degradation/persistence of pesticides

Table 4 Classification of saline, sodic, and saline/sodic soils

Classification	*Conductivity (mmhos cm^{-1})*	*Soil pH*	*Exchangeable sodium percentage*	*Soil physical condition*
Saline	>4.0	<8.5	<15	Normal
Sodic	<4.0	>8.5	>15	Poor
Saline/sodic	>4.0	<8.5	>15	Normal

pH caused by the presence of $CaCO_3$ can reduce availability of P and several micronutrients.

Saline and sodic soils are characterized by their electrical conductivity (EC_{se}), soil pH, and exchangeable Na% (Table 4). When exchangeable Na $>$ 15%, soil aggregates disperse, reducing permeability to water. In saline soils soluble salt concentration interferes with plant growth, although salt tolerance varies with plant species. In sodic soils, excess Na disperses clays and can create nutritional disorders. Saline-sodic soils exhibit both high salt and Na content.

Table 5 Typical C:N ratios for selected of plant residues, soil microbes, and soil OM

Organic substances	*C:N ratio*
Soil microorganisms	8:1
Soil organic matter	10:1
Sweet clover (young)	12:1
Barnyard manure (rotted)	20:1
Clover residues	23:1
Green rye	36:1
Corn/sorghum stover	60:1
Small grain straw	80:1
Oak	200:1
Pine	286:1
Spruce, fir	>1000:1

Specific Nutrients

Nitrogen

Most soils do not supply sufficient N to optimize productivity of nonlegume crops. Thus, supplemental N must be provided. Understanding N cycling is necessary to optimize crop yield and minimize the environmental impact of N inputs. Most soil N originated from atmospheric N_2; however, plants cannot directly metabolize N_2 into protein. Atmospheric N_2 is transformed to plant-available N by: (1) symbiotic microorganisms associated with legume roots; (2) nonsymbiotic soil microorganisms on the soil surface; (3) inorganic N in rainfall caused by lightning; and (4) manufacture of N fertilizers (from N_2).

Rainfall contains NH_4^+, NO_3^-, NO_2^-, and organic N. Atmospheric electrical discharges and industrial waste gases annually account for 1–50 kg N ha^{-1}. Symbiotic bacteria (rhizobia) specific to each host species convert N_2 to NH_4^+ in legume root nodules. N fixation by legumes ranges from 10 to 200 kg ha^{-1} year^{-1}, depending on the legume, and many soil, crop, and environmental factors. Low soil pH reduces rhizobia function through Al^{3+} toxicity and/or low Ca^{2+} supply. High soil N availability also reduces N_2 fixation. Management practices that reduce legume yield will reduce N_2 fixation. For example, enhanced water, nutrient, weed, insect, and diseases stress will reduce N_2 fixation. Poor harvest management of forage legumes that include excessive harvest frequency and premature or delayed harvest will reduce legume stand and N_2 fixation. Some nonsymbiotic bacteria and blue-green algae also fix small amounts of N_2 (1–5 kg N ha^{-1} year^{-1}) but contribute little to N availability. Industrial N fixation is the major N source. The Haber-Bosch reaction converts N_2 and H_2 to NH_3 gas used directly (anhydrous NH_3 fertilizer) or manufactured into fertilizer N products.

N mineralization/immobilization Soil comprises about 5% inorganic N and 95% inorganic N. Ammonium (NH_4^+), nitrate (NO_3^-), and to a minor extent nitrite (NO_2^-) are in soil solution and absorbed by plants. Plant-available NH_4^+ and NO_3^- depend on N applied as fertilizers or waste materials, and N mineralized from soil OM. Mineralization is a two-step process (aminization and ammonification) that converts organic N to NH_4^+. N mineralization increases with soil OM content, temperature, moisture, and O_2. During the growing season 1–4% organic N is mineralized. N immobilization is the microbial conversion of NH_4^+ and NO_3^-, to organic N. As soil microorganisms decompose plant residues, N immobilization can substantially lower NH_4^+ and NO_3^-, resulting in potential crop N deficiency. N input rates are established to satisfy N demand by the crop and soil microbes.

Relative N mineralization or immobilization depends on the plant residue C:N ratio. Soil microbes have a low (8:1) C:N ratio relative to plant residues (Table 5). Mineralization occurs with plant residues $\leq$20:1 C:N because residue N is sufficient to meet the microbial demand during decomposition. Immobilization occurs with plant residues $\geq$30:1 C:N. For ratios between 20 and 30, no net N release or immobilization occurs. Residue N content also indicates potential N immobilization/mineralization. With >2% N content, mineralization usually occurs. The residue decomposition rate depends on the quantity of residue added, soil inorganic N content, temperature, moisture, and residue characteristics (particle size and lignin/waxes content). Uncultivated soils have a relatively constant amount of plant residue returned to the soil and usually low N mineralization. When tilled, N mineralization increases, reducing soil OM content and ultimately N availability. Any change in OM content dramatically reduces the quantity of N mineralized and thus soil N availability to crops.

Nitrification The NH_4^+ produced from mineralization or other N inputs is converted to NO_3^- through a two-step nitrification reaction: $NH_4^+ \rightarrow NO_2^- \rightarrow NO_3^-$. The NO_3^- produced is mobile and subject to potential leaching losses. Knowledge of the factors affecting nitrification and careful management of N inputs can reduce N leaching losses. Generally the environmental conditions that favor plant growth are optimum for nitrification (Table 6).

NH_4^+ fixation Vermiculite and mica clay minerals can fix NH_4^+ in interlayer spaces. Some fixed NH_4^+ can be replaced with Ca and Mg. The availability of fixed NH_4^+ ranges from negligible to relatively high. NH_4^+ fixation is important in certain soils, fixing nearly 30% of applied N.

Denitrification Under low O_2 conditions, anaerobic organisms obtain O_2 by converting NO_2^- and NO_3^- to N_2, N_2O, NO, and NO_2 gases. Denitrifying organisms also exist in well-oxygenated soils mostly near roots (rhizosphere C source) and in anaerobic microsites or soil aggregates where O_2 diffusion is reduced. Denitrification is enhanced by plants because of readily available C in root exudates and sloughed-off root tissues. Denitrification potential is high in most soils; however, the quantity of gaseous N loss varies due to fluctuations in environmental conditions (H_2O content, temperature, etc.) and other factors (Table 6). Denitrification occurs primarily during periods of high precipitation (heavy rains, irrigation, snow melt), and continuously as small losses from anaerobic microsites in soil aggregates. Under denitrifying conditions, NO_2^- can accumulate, which is toxic to plants. High rates of band-applied N fertilizers elevate NH_4^+ levels and pH, enhancing NO_2^- accumulation. Denitrification is a useful remediation tool for reducing excessive NO_3^- in soil water. The use of riparian buffers to reduce NO_3^- in groundwater is well established.

Volatilization Compared with NH_3 volatilization associated with urea or NH_4^+ fertilizers, NH_3 loss from N mineralized from OM is small. Volatilization of NH_3 depends on high NH_4^+ in soil solution and high soil pH given by $NH_4^+ \rightarrow NH_3 + H^+$. The quantity of NH_3 loss depends on the interaction of many soil, environmental, and N fertilizer factors, but is highest in coarse-textured, calcareous soils, and with surface crop residues (Table 6).

Phosphorus

Availability of P depends on solution P concentration and the ability of soil P fractions to resupply or buffer solution P. As solution P decreases with plant uptake,

Table 6 Factors affecting nitrification, denitrification, and volatilization of soil N

Factor	*Comment*
Nitrification	
Supply of NH_4^+	Conditions favoring NH_4^+ production (mineralization) increase nitrification
Soil pH	Wide pH range, optimum pH is 8.5. Nitrifying bacteria need adequate Ca and P
Soil aeration	Nitrifying bacteria require O_2; good soil structure is essential to gas exchange/aeration
Soil moisture	Nitrification is highest at field capacity soil water content
Soil temperature	Optimum temperature 25–35°C. To minimize N leaching losses, soil temperature <10°C to fall apply N
Denitrification	
Decomposable organic matter	Fresh crop residues (C source) stimulate microbial growth and respiration
Soil moisture	High H_2O content impedes O_2 diffusion, increasing anaerobic microbe activity
Soil pH	Denitrifying bacteria prefer pH > 5.0. N gas released is also influenced by pH
Soil temperature	Denitrification increases rapidly from 2 to 25°C, decreasing above 60°C
NO_3^- supply	Increasing NO_3^- increases denitrification potential
Volatilization	
Soil pH	NH_3 loss increases with increasing pH
Buffer capacity	High BC decreases NH_3 loss by increased adsorption and pH buffering
N placement	Volatilization losses are greater with broadcast versus banded N fertilizers
Soil temperature	Volatilization increases with increasing temperature up to about 45°C
Soil moisture	Maximum NH_3 loss occurs at field capacity and during slow drying conditions
Crop residue	Residues increase NH_3 through wet, humid conditions at soil surface

P minerals dissolve, P adsorbed on clay mineral surfaces desorb, and organic P mineralizes to resupply solution P. Maintaining adequate plant-available P in solution (intensity) depends on P desorption and dissolution reactions (quantity) that resupply solution P taken up by plants (Figure 6). The larger the BC, the greater the ability to buffer solution P. Plants absorb $H_2PO_4^-$ and HPO_4^{2-}. Average solution P concentration is 0.05 ppm and varies widely among soils. Solution P required by plants varies from 0.003 to 0.3 ppm that is replenished nearly 100 times during the growing season. Rapid replenishment of solution P is essential near actively absorbing roots.

Organic P About 50% of total soil P is organic and occurs primarily as inositol phosphates, phospholipids, and nucleic acids. P mineralization represents microbial degradation of organic P in plant/animal residues and soil OM to release $H_2PO_4^-/HPO_4^{2-}$. P mineralization increases with increasing OM. The C/P ratio or P content of decomposing residues regulates whether P mineralization or immobilization occurs (Table 7). P and N mineralization/immobilization processes are similar. Inorganic mineralized P can be removed from solution through adsorption to mineral surfaces and precipitation as Al-, Fe-, or Ca-P minerals. The quantity of P mineralized varies widely but contributes about 3–10 kg ha^{-1} in temperate region soils.

Inorganic P Inorganic P in solution can be: (1) absorbed by plants; (2) immobilized by microorganisms; (3) adsorbed to mineral surfaces; and/or (4) precipitated as secondary P compounds. Adsorption and precipitation reactions are referred to as P fixation and are a continuous sequence of precipitation and adsorption. In acid soils, inorganic P precipitates as Fe-P or Al-P minerals and/or is adsorbed to Fe/Al oxide surfaces. In neutral or high-pH soils, Ca-P minerals precipitate and/or P is adsorbed to clay and $CaCO_3$. Adsorption dominates under low solution P, while precipitation occurs when solution P levels exceed a mineral's solubility product (K_{SP}). Ultimately, solution P concentration is controlled by the solubility of inorganic P minerals. For example, decreasing solution $H_2PO_4^-$ by plant uptake causes variscite ($AlPO_4 \cdot 2H_2O$) to dissolve and resupply solution P according to:

$$AlPO_4 \cdot 2H_2O + H_2O \leftrightarrow H_2PO_4^- + H^+ + Al(OH)_3$$

Many factors affect P fixation in soils that ultimately influence P availability to plants (Table 8). P fixation is strongly influenced by soil pH, where P availability is maximum near pH 6.5. At low pH values, fixation occurs largely through adsorption to Fe/Al oxides and precipitation as $AlPO_4$ and $FePO_4$. As pH increases, Fe and Al concentration decreases, reducing P adsorption/precipitation. At pH >7.0, Ca-P minerals precipitate and P is adsorbed to $CaCO_3$, causing solution P to decrease. Thus, minimum P adsorption occurs at pH 6.0–6.5.

In soils with high P fixation capacity (high clay content or BC), solution P will decrease quickly after P fertilization. With low P fixation capacity solution P can be elevated for several months or years depending on P rate and application method. With high P fixation, fertilizer P should be applied in smaller, more frequent applications. When fertilizer P is broadcast-applied, P is in contact with more soil, increasing P fixation, compared to band-applied P. Band placement decreases contact between soil and fertilizer, reducing P fixation and increasing plant-available P.

Table 7 Influence of residue C:P on P mineralization and/or immobilization

C:P ratio	*%P in residue*	*Mineralization/immobilization*
< 200	> 0.3	Net mineralization of organic P
200–300	0.2–0.3	No gain or loss of inorganic P
>300	< 0.2	Net immobilization of inorganic P

Potassium

Soil K content ranges between 0.5 and 2.5% and is substantially greater than that absorbed by plants. Depending on the soil, <10% soil K is plant-available. Exchangeable and solution K provides plant-available K during the season and nonexchangeable K resupplies solution and exchangeable K. Nonexchangeable

Table 8 Factors affecting P fixation in soils

Factor	*Comment*
Soil minerals	Fe/Al oxides can adsorb large amounts of solution P. P adsorption is greater with 1:1 clays than 2:1 clays. P adsorption increases with increasing clay content
Soil texture	Increasing clay content increases P adsorption
Soil pH	P adsorption to Fe/Al oxides decreases with increasing pH
Soil organic matter	Organic matter can 'coat' minerals, reducing P adsorption
Soil temperature	P adsorption increases with increasing temperature

K release is slow and essentially unavailable to crops during a season. Unlike N, P, and other nutrients, K availability is not influenced by soil OM.

Solution K averages about 5 ppm and is in equilibrium with exchangeable K. Equilibrium between solution and exchangeable K is represented by the *Q/I* ratio described for P. *Q/I* is directly proportional to CEC and measures relative ability to resupply solution K^+. High *Q/I* represents a high K buffer. Adding fertilizer K will increase *Q/I* and K availability. Low *Q/I* represents low CEC, sandy soils exhibiting low K buffer. The ability of soil to buffer solution K^+ as plants remove solution K^+ is partially controlled by the release rate of interlayer K. About 90% total soil K is nonexchangeable or mineral K. Nonexchangeable K resupplies exchangeable and solution K^+ as they are reduced during plant uptake. The K release rate is controlled through weathering of K minerals (micas > feldspars). Low solution K encourages interlayer K release. Mica and vermiculite minerals can 'fix' K, which represents solution K^+ entering the interlayer region. NH_4^+ ions can also be fixed. Fixation of K^+ and NH_4^+ occurs more in clay soils. Fixation of K reduces solution K, but can be released again to available K. Many factors influence K availability (Table 9).

Sulfur

Like N, sulfur (S) occurs in organic (≥90%) and inorganic (SO_4^{2-}) forms. Although large temporal variation occurs, solution S ranges from 1 to 20 ppm SO_4^{2-}, where sandy soils often contain ≤4 ppm. Near industrial areas, solution SO_4^{2-} can be increased through deposition caused by SO_2 pollution. Like NO_3^-, SO_4^{2-} can be leached, especially in sandy soils with a high quantity of percolating water. Adsorbed SO_4^{2-} can contribute to plant-available S in acid soils with relatively high AEC and/or Al/Fe oxide content. Many soil factors affect SO_4^{2-} adsorption (Table 10). The amount and type of clay and soil pH have the greatest influence on the quantity of SO_4^{2-} adsorbed. The availability of S depends mostly on mineralization of organic S to SO_4^{2-}. S mineralization ranges from 5 to 15 kg SO_4-S ha^{-1} $year^{-1}$. Immobilization converts SO_4^{2-} to organic S. Factors that affect microbial activity influence S mineralization/immobilization

Table 9 Factors affecting K availability to crops

Factor	*Comment*
Clay minerals	Increasing 2:1 clay content increases K availability
CEC	Increasing CEC increases K availability
Exchangeable K	Increasing exchangeable K increases K availability
K fixation	Soils with greater K fixation potential require greater K inputs to increase K availability
Soil moisture	Increasing solution K or H_2O content increases K diffusion and availability
Soil temperature	Increasing soil temperature increases K availability (> diffusion and K release rate)
Soil pH	At low pH K uptake can be reduced by Al toxicity. Under higher pH, > exchangeable Ca/Mg competes with K for uptake sites, reducing K uptake

CEC, cation exchange capacity.

Table 10 Factors affecting SO_4^{2-} adsorption and S mineralization/immobilization in soils

Factor	*Comment*
Adsorption	
Clay content	Adsorption increases with clay content. SO_4^{2-} adsorption follows kaolinite > illite > montmorillonite
Hydrous oxides	Fe/Al oxides contribute most of the SO_4^{2-} adsorption in severely acid soils
Soil depth	SO_4^{2-} adsorption is greater in subsoils due to greater clay and Fe/Al oxides
Soil pH	AEC increases with decreasing pH. Little SO_4^{2-} adsorption at pH >6.5
SO_4^{2-} concentration	Adsorbed SO_4^{2-} increases with solution SO_4^{2-}
Other anions	Adsorption strength decreases according to $OH^- > H_2PO_4^- > SO_4^{2-} > NO_3^- > Cl^-$
Soil OM	AEC site on OM can contribute to SO_4^{2-} adsorption.
Mineralization and immobilization	
S content of OM	S mineralized increases with increasing S content in residue. C/S ratio ≤ 200 S mineralization occurs; C/S = 200–400 no net mineralization occurs; C/S ≥ 400 S immobilization occurs
Soil temperature	S mineralization is optimum between 20 and 40°C
Soil moisture	S mineralization is optimum at 60% of field capacity

OM, organic matter; AEC, anion exchange capacity.

(Table 10). Soils with low OM (<1.2–1.5%) and/or course texture are often S-deficient.

Calcium and Magnesium

In acid and neutral pH soils, Ca and Mg occur on exchange sites and in minerals. In calcareous soils Ca/Mg also exist as $CaCO_3$, $MgCO_3$, and/or $CaMg(CO_3)_2$. As exchangeable cations, they behave similarly to K. As Ca/Mg in solution decreases with plant uptake or leaching, Ca/Mg desorbs from CEC to resupply the solution (Figure 2). Plant uptake of Ca/Mg depends mostly on the amount of solution Ca/Mg, soil pH, and %Ca/Mg on CEC. Solution Ca/Mg decreases by: (1) plant uptake; (2) leaching; (3) adsorbed on CEC; and/or (4) precipitated as minerals (in arid climates).

Solution Ca and Mg ranges between 5 and 300 ppm Ca and 5 and 50 ppm Mg (15 ppm Ca and 10 ppm Mg is adequate for most crops), with the higher levels in arid regions. Sandy, humid-region soils are generally low in Ca and/or Mg, whereas clay soils are usually higher in exchangeable Ca/Mg. Low-Ca soils require supplemental Ca, usually through liming ($CaCO_3$), to increase pH. Many crops respond to Ca and Mg when Ca <25% and Mg <10% of CEC. High exchangeable K can interfere with Mg uptake. High NH_4^+ levels from fertilization can induce Mg deficiency in low-Mg soils.

Table 11 Factors affecting micronutrient availability to crops

Factor	*Comment*
Iron	
Soil pH	Deficiencies most often observed on high-pH (>7.2), calcareous, arid-region soils
Poor aeration	Increase in CO_2 in wet, poorly drained soils increases HCO_3^-, enhancing Fe stress
Soil OM	Low soil OM can enhance Fe stress due to low chelation, especially on high-pH soils
Nutrient interactions	High Cu, Mn, Zn, Mo, and P can reduce Fe uptake
Plant species	Plant species and varieties vary in susceptibility to Fe stress
Zinc	
Soil pH	Deficiencies most often observed on high-pH soils
Soil OM	Low soil OM can enhance Zn stress due to low chelation, especially on high-pH soils
Nutrient interactions	High Fe, Cu, Mn, and P can reduce Zn uptake
Climate	Zn deficiencies are more pronounced during cool, wet seasons
Plant species	Plant species and varieties vary in susceptibility to Zn stress
Copper	
Soil texture	Solution Cu is lower in leached and calcareous sands
Soil pH	Deficiencies most often observed on high-pH soils
Soil OM	Cu strongly complexes with OM. High OM can reduce Cu availability
Nutrient interactions	High Fe, Zn, and P can reduce Cu uptake
Climate	Cu deficiencies are more pronounced during cool, wet seasons
Plant species	Plant species and varieties vary in susceptibility to Cu stress
Manganese	
Soil pH	Deficiencies most often observed on high-pH, calcareous, coarse-textured soils
Poor aeration	Mn availability can be increased by wet, poorly aerated soil
Soil OM	Mn availability increases with OM. Very high OM can reduce Mn availability
Nutrient interactions	High Cu, Fe, or Zn can reduce Cu uptake
Climate	Mn deficiencies are more pronounced during cool, wet seasons
Plant species	Plant species and varieties vary in susceptibility to Mn stress
Boron	
Soil pH	B availability decreases with increasing pH, especially at pH >6.5
Soil moisture	B deficiency is enhanced with low soil moisture conditions
Soil OM	Increasing OM increases B availability
Soil texture	Coarse-textured, well-drained, sandy soils are often low in B
Nutrient interactions	Low Ca increases B uptake and potential toxicity
Plant species	Plant species and varieties vary in susceptibility to B deficiency and toxicity
Molybdenum	
Soil pH	Mo availability increases with increasing pH
Fe/Al oxides	High Fe/Al oxides increase Mo adsorption (decrease available Mo)
Nutrient interactions	P enhances Mo absorption by plants. S, Cu, Mn can decrease Mo uptake
Climate	Mo deficiency is more severe under dry soil conditions
Plant species	Plant species and varieties vary in susceptibility to Mo deficiency

OM, organic matter.

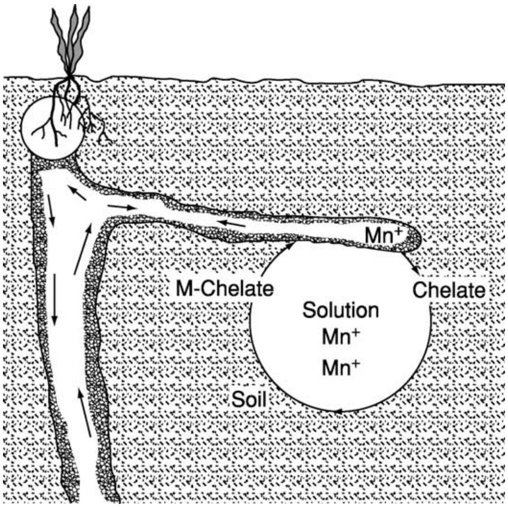


Figure 8 Cycling of chelate-bound micronutrients in soil solution to provide plant-available micronutrients. Chelates in solution bind micronutrients to metal cations and transport them to the root surface where micronutrient is released to plant uptake.

Micronutrients

Micronutrients occur in soil as: (1) primary and secondary minerals; (2) adsorbed to mineral and OM surfaces; (3) constituents of OM; and (4) solution (Figure 2). The proportion of each depends on the micronutrient, but all impact plant availability.

Iron, zinc, manganese, copper Plant-available Fe, Zn, Mn, and Cu are controlled by mineral solubility and soil OM. The solubility of Fe, Zn, Mn, and Cu minerals is generally very low and decreases with increasing pH. In most soils, solution Fe is not sufficient to meet plant need, thus another process increases Fe availability since most soils (except high-pH, calcareous soils) provide adequate Fe for plant growth. Soluble organic compounds (chelates) in soil solution complex micronutrients and increase solution Fe, Zn, Mn, and Cu concentration and diffusion to plant roots (Figure 8). Increasing soil OM can improve micronutrient availability. Factors influencing micronutrient availability are illustrated in Table 11.

Boron, chloride, molybdenum, nickel Boron (B) occurs in solution ($H_3BO_3^0$), minerals, adsorbed on clays and Fe/Al oxides, and in OM. Adsorbed and mineral B buffer solution B. Increasing pH, clay content, and OM favor B adsorption. Soil OM complexes with B and increases B availability. Most of the chloride (Cl^-) exists in soil solution, with very little occurring in mineral, organic, or adsorbed forms. In arid-region soils, Cl salts can precipitate and accumulate in surface soil, causing potential salinity and sodicity problems. Low-Cl soils can enhance disease incidence on some crops. Excess Cl in irrigation water can reduce yield of salt-sensitive crops. Molybdenum (Mo) occurs in minerals, adsorbed on Fe/Al oxides, complexed in OM, and in soil solution (MoO_4^{2-}). Solution Mo increases with increasing soil pH. Plant-available Ni is provided by adsorbed and mineral dissolution. Factors affecting B, Cl, Mo, Ni availability are shown in Table 11.

See also: **Fertilizers and Fertilization**; **Nutrient Availability**; **Nutrient Management**; **Quality of Soil**

Further Reading

Adams F (ed.) (1984) *Soil Acidity and Liming*. Madison, WI: Soil Science Society of America.

Barber SA (1984) *Soil Nutrient Bioavailability: A Mechanistic Approach*. New York: John Wiley.

Bresler E, McNeal BL, and Carter DL (1982) *Saline and Sodic Soils*. New York: Springer-Verlag.

Follett RF (ed.) (1989) Nitrogen management and groundwater protection. In: *Developments in Agricultural Managed-Forest Ecology #21*. New York: Elsevier.

Khasawneh FE (ed.) (1980) *The Role of Phosphorus in Agriculture*. Madison, WI: American Society of Agronomy, Crop Science Society of America, Soil Science Society of America.

Lindsay WL (1979) *Chemical Equilibria in Soils*. New York: John Wiley.

Mengel K and Kirkby EA (1987) *Principles of Plant Nutrition*. Bern, Switzerland: International Potash Institute.

Mortvedt JJ, Giordano PM, and Lindsay WL (eds) (1991) *Micronutrients in Agriculture, No. 4*. Madison, WI: Soil Science Society of America.

Mortvedt JJ, Murphy LS, and Follett RH (1999) *Fertilizer Technology and Application*. Willoughby, OH: Meister.

Munson RD (ed.) (1985) *Potassium in Agriculture*. Madison, WI: Soil Science Society of America.

Rendig VV and Taylor HM (1989) *Principles of Soil–Plant Relationships*. New York: McGraw-Hill.

Stevenson FJ (ed.) (1982) *Nitrogen in Agricultural Soils*. Madison, WI: American Society of Agronomy.

Stevenson FJ (1986) *Cycles of Soil*. New York: John Wiley.

Tabatabai MA (ed.) (1986) *Sulfur in Agriculture. No. 27*. Madison, WI: ASA, CSSA, Soil Science Society of America.

FERTILIZERS AND FERTILIZATION

H W Scherer, University of Bonn, Bonn, Germany

Introduction

The main nutrient supplies for plants are released from soil reserves, decomposing biological residues (leaves, roots, etc.), aerial deposition, biological nitrogen fixation and, in the case of crop plants, fertilizers. When considering the availability of elements for plants, soil factors such as cation exchange capacity, acidity, redox potential, temperature, synergistic and antagonistic effects of other elements must be taken into account. The most important sources of elements in soils are the parent materials from which they are derived and soil provides the principal natural source of nutrients, while fertilizers provide the additional input to increase crop yields.

Total world consumption of fertilizer nitrogen (N), phosphorus (P), and potassium (K), in 1999/2000 was 85, 15, and 19 million t year^{-1}, respectively. In the future the demand for fertilizers is expected to increase with world population growth and the need for food. The projected increase in fertilizer consumption will mainly occur in developing countries, where food production urgently needs to be increased.

Nitrogen

Soil Nitrogen

N is the mineral nutrient most in demand by plants. Inorganic soil N mainly occurs as ammonium (NH_4^+) and nitrate (NO_3^-). In the surface layer of most soils the share of inorganic N of total soil N is usually less than 5%, while over 90% of N is organically bound. Most of the N in soil organic matter is in the form of stable humic compounds, while a minor part, mainly amino acids and amino sugars, is labile with a rapid turnover. Such compounds represent prime organic substrates in deamination reactions, where the removal of NH_4^+ is mostly carried out by enzymes. Once NH_4^+ is formed by mineralization it can be taken up by plants. However, mineralization depends on different factors that cannot be controlled and the release of NH_4^+ from organic N compounds is not necessarily timed to fit the plant's need. In most soils NH_4^+ may be fixed by clay minerals or can be held on negatively charged soil complexes or humic compounds. This reduces its mobility and therefore the availability for plants as compared to the more mobile NO_3^- ion, which is formed by nitrification of NH_4^+. Nitrification is important from the standpoint of soil fertility.

N Uptake of Plants

Plants take up N as NO_3^-, which is the most abundant form of inorganic N in the soil solution, and as NH_4^+. The uptake of both N forms depends mainly on their availability. In many crops the combination of NO_3^- and NH_4^+ results in better growth than when either form is applied alone: optimal proportions depend strongly on the total concentrations supplied. Whether NO_3^- or NH_4^+ as sole nitrogen source is better for plant growth depends on different factors. One important factor is the plant species. Plants that are adapted to acid soils or to low soil redox potential have a preference for NH_4^+, while plants with a preference for high-pH soils prefer NO_3^-.

Soil Test

In intensive agriculture deciding the optimal N fertilizer rate is a major challenge from an economic as well as ecological point of view. The optimal N fertilizer rate is the difference between the N demand of the crop and the N supply from different sources. Therefore, in predicting N fertilizer needs, in addition to the N demand of the crop, both the residual mineral N as well as the mineralizable N must be taken into consideration. The importance of determining residual mineral N as a basis for N fertilizer recommendation is already well established. Especially in the deeper soils of cooler, drier climates the amount of mineral N must be taken into account. Renewed interest in developing improved methods, taking the N mineralization capacity into account, has mainly been stimulated by concern over the environmental consequences of applying more fertilizer N than is demanded by the crop. However, the real crop need for N is not known before harvest. Therefore, yield is forecast using recent average yield data from the local area. Because an exact yield estimate is not possible, the crop has to be observed during the growing period and the N fertilizer plan adjusted. Depending on the N fertilizer form applied, N is more or less rapidly available to plants. However, calculating the optimum N application rate, it must be considered that even under optimal conditions N fertilizers are never fully utilized by the crop.

Time of N Fertilizer Application

Depending on soil and especially on climatic conditions, splitting the total N application rate into two or even more increments is recommended for maximum N fertilizer efficiency, whereby the first N fertilizer application is recommended shortly before the period of highest demand and of most rapid plant growth. The time of sidedressing depends on the crop. For cereals, maximum N fertilizer efficiency is reached by the latest possible application which is compatible with the development stage that permits ready uptake. However, for crops like sugar beet with a high N demand at the early growing stage for developing an extensive leaf surface, a late N fertilizer application is not desirable. To assure high sugar concentrations in the roots, sugar beets should be under conditions of mild N deficiency in the end of the growing period.

N Fertilizers

Most N fertilizers contain NO_3^- and NH_4^+ as the N carrier (Table 1). Caused by the sorption of NH_4^+ on soil colloids, the uptake rate of NO_3^- is normally higher under field conditions and therefore plants respond faster to NO_3^- application. However, N losses by leaching or denitrification may be higher with NO_3^- application compared with NH_4^+ application. Especially on paddy soils, denitrification losses may be high. For this reason fertilizers containing NH_4^+ or urea are used on these soils. On the other hand, NH_3 losses may occur in alkaline soils, when NH_4^+-containing fertilizers are applied, caused by the conversion of NH_4^+ to NH_3.

Table 1 Nitrogen (N) fertilizers

Fertilizer	*Chemical composition*	*N concentration (%)*
Ammonium nitrate	NH_4NO_3	35
Nitrochalk	$NH_4NO_3 + CaCO_3$	27
Ammonium nitrate sulfate	$(NH_4)_2SO_4 \cdot NH_4NO_3$	26
Ammonium sulfate	$(NH_4)_2SO_4$	21
Aqueous ammonia	NH_3, NH_4OH	>10
Anhydrous ammonia	NH_3	82
Urea	$CO(NH_2)_2$	46
Calcium cyanamide	$CaCN_2$	22

Adapted from Mengel K and Kirkby EA (2001) *Principles of Plant Nutrition.* Dordrecht: Kluwer Academic, with permission.

Ammonium nitrate, which is an explosive salt, is prohibited in some countries for direct use as an N fertilizer, but is often used to prepare liquid fertilizers. To reduce the danger of fire, limestone is added to ammonium nitrate, resulting in so-called nitrochalk. Neutralization of nitric and sulfuric acid with NH_3 results in the formation of the double salt of ammonium sulfate and ammonium nitrate (Figure 1).

Ammonium nitrate sulfate, a double salt of NH_4NO_3 and $(NH_4)_2SO_4$ in equal proportions, is formerly known as leunasalpeter. It is widely used on forage and grass seed crops and on autumn-seeded small grains. (Potassium nitrate is often used for foliar application.) Some years ago ammonium sulfate, which results in acidification of the soil, was an important fertilizer, but its consumption has declined in recent years and it was substituted by urea, which has a higher N content. Urea is a major N fertilizer. It is converted to NH_4^+ by urease in the soil. This rapid hydroloysis may result in significant N losses through NH_3 volatilization. Volatile losses of NH_3 originating from urea may be reduced by applying urease inhibitors that delay the formation of NH_4^+.

Calcium cyanamide is a slow-reacting N fertilizer, which contains N in the amide and cyanide form. This N fertilizer must be converted to urea, which is split into NH_3 and CO_2 before plant uptake. During conversion in the soil intermediate toxic products are formed, which are nitrification inhibitors as well as herbicides. For this reason the application of calcium cyanamide is recommended before sowing or transplanting.

Anhydrous ammonia is a liquid N fertilizer, which contains NH_3 in a liquid under pressure. It is a high-grade N fertilizer and therefore of considerable

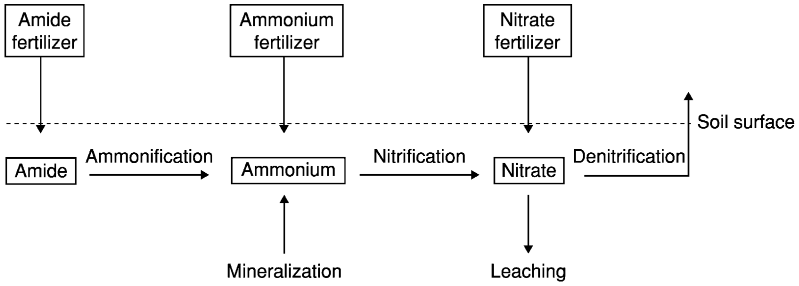


Figure 1 Turnover of nitrogen fertilizers in the soil.

advantage in terms of transport costs. However, its use is often restricted, because the liquid is under pressure, and therefore it needs special safety precautions and equipment for transportation and application. To minimize NH_3 losses by volatilization an injection assembly is used to apply it to the soil.

Aqueous ammonia contains about 25% NH_3 and the solution is only under low pressure. Therefore the handling of this solution is less complicated compared with anhydrous ammonia. NH_3 losses can usually be avoided by applying it at sufficient depth and under appropriate soil physical and moisture conditions.

Slow-Release and Bioinhibitor-Amended Nitrogen Fertilizers

Only 50–60% of fertilizer N added to the soil is usually taken up by the crop during the season of application. An approach to increase the efficiency of N fertilizers is to control the rate of N fertilizer dissolution. Controlled-release fertilizers, which are expensive to produce and therefore mainly used in the production of high-value crops, can be divided into two groups: coated fertilizers and uncoated fertilizers with a low solubility. The aim of using controlled-release fertilizers is to provide nutrients according to crop demand.

Coated fertilizers Different types of coating are used: impermeable coatings with tiny holes through which solubilized materials diffuse, and semipermeable coatings through which water diffuses until the internal osmotic pressure ruptures the coating or impermeable coatings that must be broken, for example, by biological actions. Further coatings may function only as physical barriers or be a source of plant nutrient. The most important coating materials, which are usually applied in intimate contact with the surface of the fertilizer particle are waxes, polymers, and sulfur.

Osmocotes are surrounded by a plastic shell, which allows water to diffuse into the shell, which tears as a result of the uptake of water and the nutrients diffuse into the soil. With sulfur-coated urea (SCU), water vapor transfers through the sulfur coating, solubilizes the urea within the shell, builds up sufficient osmotic pressure to disrupt the coating, and urea solution is released. The decomposition of the sulfur coat also depends on the oxidation of the sulfur, which is mainly brought about by soil microorganisms. Hence the availability of the sulfur-coated fertilizers depends on microbial activity.

Because the release of nutrients from the fertilizer depends on soil water content, and plant growth depends on soil moisture, nutrient release may also be adapted to plant growth.

Uncoated fertilizers While urea is readily soluble in water and quickly decomposed to release NH_4^+, it forms several chemical reaction products that are useful as slow-release N fertilizers. Urea reacts with several aldehydes to form compounds that are sparingly soluble in water. Commercial products are ureaform (UF), crotonylidendiurea (CDU), and isobutylidenurea (IBDU). After application to the soil, these fertilizers are prone to hydrolysis and release urea. The rate of microbiological decomposition is controlled by the property of the aldehyde as well as by soil temperature.

Nitrification Inhibitors

A controlled supply of N may, presumably, also be achieved by the use of any ammonium source in combination with a nitrification inhibitor (DCD = dicyandiamide; N-serve = nitrapyrin), which blocks the oxidation of NH_4^+ to NO_2^- (Figure 2). The loss of N by leaching or denitrification is therefore reduced. The amount of nitrification inhibitor needed to reduce the rate of nitrification can be reduced when it is granulated with the ammonium fertilizer.

Methods of N Application

Most the N fertilizer is applied by surface broadcasting, followed by injection of liquids and gases, incorporation by attachment to plow, planter, or cultivator, and distribution in irrigation water.

Especially in orchard or sod conditions where immediate fertilizer incorporation is not possible, surface broadcasting has some shortcomings. Under dry conditions on soils with a high pH, significant amounts of N can be lost by NH_3 volatilization when N is applied as urea or as NH_4^+ salts.

The injection of liquids and gases requires careful adaptation to the conditions of soil water. If the soil is too wet or too dry, volatilization losses may occur. Even under optimum soil water conditions injection must be at sufficient depth.

By application with the irrigation water (fertigation), N is applied during the whole growing season according to crop needs. A disadvantage associated with sprinkler irrigation is the chance of nonuniform N distribution caused by excessive wind.

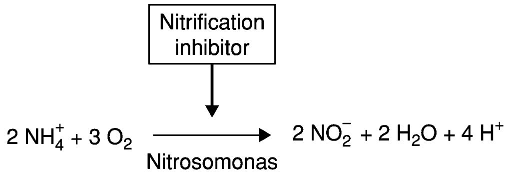


Figure 2 Inhibition of nitrification by a nitrification inhibitor.

Phosphorus

Soil Phosphorus

P in soils is present in the soil solution (plant-available), as labile phosphate precipitates and adsorbed to soil particles, mainly clay minerals (potentially available to plants), as nonlabile phosphate in the form of calcium, iron, and aluminum phosphate (not plant-available), in organic form, including P in soil organic matter (released after mineralization), and in living soil biomass (Figure 3). The quantity of P in soil solution in the upper soil layer is in the range of 0.3–3 kg P ha^{-1} and must be replenished by mobilization of phosphate from the labile pool within a time span of hours or days. There is a critical limit of readily soluble soil phosphate below which yields become severely constrained. For a high production level as well as for sustainable agriculture the maintenance of the readily soluble soil phosphate pool is a prerequisite.

Soil Test

In order to recommend an appropriate P fertilizer application, a soil test is indispensable for both soils with insufficient as well as for those with high amounts of available P. However, although reliable soil tests for available phosphate are at our disposition, regular P soil testing only covers a limited percentage of agricultural land. Soil tests avoid wasting the limited raw-material phosphate and reduce the risk of P leaching.

Different extraction methods are used as an index of plant-available P in soils. However, these methods never measure the quantity of P available to a crop, but determine more or less a pool of soil P that is somehow related to that portion of soil P which is plant-available.

To determine the P concentration of the soil solution the soil is extracted with water or a diluted salt solution (i.e., 0.02 mol l^{-1} $CaCl_2$). The objective of this method is to determine the P concentration level that directly limits plant growth. The most common methods used worldwide to extract plant-available P are probably the acid fluoride extract and the alkaline bicarbonate extract. In Europe an extraction using lactate is common.

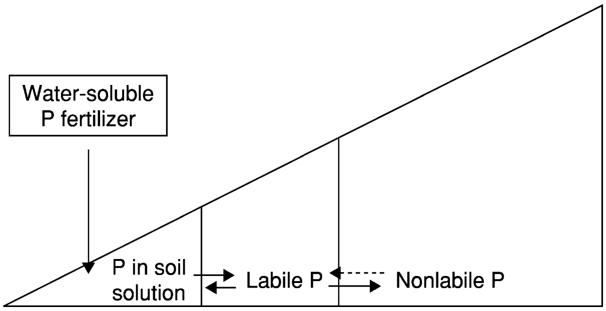


Figure 3 Phosphorus (P) dynamics in the soil. (Adapted from Mengel K and Kirkby EA (2001) *Principles of Plant Nutrition*. Dordrecht: Kluwer Academic, with permission.)

Saving P is also a question of an efficient use of soil and fertilizer P. Agronomic measures for improving phosphate efficiency are fertilizing according to the results of soil tests, providing optimum pH conditions for phosphate availability, using crop species in the crop rotation, which are able to mobilize less soluble soil phosphates, and applying appropriate P fertilizers.

P Application Rates

According to available soil P and crop species, P application rates generally range from 20 to 80 kg P ha^{-1}. As compared with slow-growing crops, crops with a high growth rate, for example corn, alfalfa, or sugar beet, require higher amounts of available P. On soils with a high P fixation capacity, up to 200 kg P ha^{-1} is recommended.

To avoid a decline in soil fertility, phosphate that has been removed by crops or is lost needs to be replenished by P fertilizer application. Less than 15% of P fertilizer applied to a soil is normally taken up by the crop grown immediately after application, while the rest is converted to soil phosphate and the recovery becomes less in the following years.

At neutral soil pH, fertilizer phosphate added to soils will at first be adsorbed or precipitated in more-soluble calcium phosphate forms. Later the phosphate of these forms will be shifted to apatite. In more acid soils rich in sesquioxides, adsorption is the strongest phosphate sink. Besides adsorption, phosphate occlusions occur in soils with a low pH.

P Fertilizers

The raw material for the production of P fertilizers is rock phosphate (apatite). Depending on their provenance, rock phosphates differ in fertilizer value. The solubility of apatites depends on the degree of isomorphic substitution of PO_4^{3-} by CO_3^{3-} and is higher the more phosphate is substituted by carbonate. While hard crystalline apatites are very insoluble and useless as a P fertilizer, soft rock phosphates may be used after grinding under acidic soil conditions (pH <5.5), for example in tropical regions, where soluble P fertilizers are prone to leaching. Further, organic farmers, who do not apply chemically treated rock phosphate, are allowed to use ground soft rock phosphate. Over the past decade the direct application of rock phosphate declined and at present it constitutes 0.8% of total global mineral P consumption.

Table 2 Phosphorus (P) fertilizers

Fertilizer	*Chemical composition*	*P concentration (%)*
Superphosphate	$Ca(H_2PO_4)_2 + CaSO_4$	8–10
Triple superphosphate	$Ca(H_2PO_4)_2$	20
Monoammonium phosphate	$NH_4H_2PO_4$	21–22
Diammonium phosphate	$(NH_4)_2H_2PO_4$	24
Basic slag	$Ca_3P_2O_8 \cdot CaO + CaO \cdot SiO_2$	4–10
Sinterphosphate	$CaNaPO_4 \cdot Ca_2SiO_4$	11–13
Partially acidulated rock phosphate	$Ca(H_2PO_4)_2$, $CaHPO_4$, apatite	10
Ground rock phosphate	Apatite	13

Adapted from Mengel K and Kirkby EA (2001) *Principles of Plant Nutrition.* Dordrecht: Kluwer Academic.

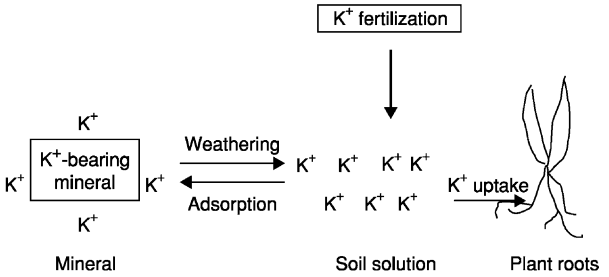


Figure 4 Potassium dynamics in the soil. (Adapted from Mengel K and Kirkby EA (2001) *Principles of Plant Nutrition.* Dordrecht: Kluwer Academic, with permission.)

Treating ground rock phosphate with sulfuric acid results in the formation of superphosphate, a mixture of $Ca(H_2PO_4)_2$ and $CaSO_4$, and treating with phosphoric acid results in the formation of triple superphosphate (Table 2). Both are water-soluble P fertilizers and suitable for most soil types. However, after the application of water-soluble P fertilizers, reactions between the phosphate and soil particles take place which reduce P availability. Under tropical conditions with a high risk of P leaching, water-soluble P fertilizers are not recommended. Partially acidulated P fertilizers are derived from rock phosphates which have been treated with amounts of sulfuric or phosphoric acids: only part of the apatite is converted to water-soluble phosphate and the rest remains as apatite. Therefore this P fertilizer contains only a part of plant-available P.

Mono- and diammonium phosphates are produced by adding NH_3 to phosphoric acid. They are often used for band application, which increases P efficiency, because less P is adsorbed by soil particles and P uptake is increased by the root attraction by NH_4^+.

Sinterphosphate is produced by disintegration of rock phosphate with Na_2CO_3 and silica at a temperature of about 1250°C and contains $CaNaPO_4$ and Ca_2SiO_4.

Basic slags ($Ca_3(PO_4) \cdot CaO + CaO \cdot SiO_2$) are a by-product of the steel industry.

Potassium

Soil Potassium

Potassium is one of the three major crop nutrients. Its main source for growing plants under natural conditions is the weathering of K^+-bearing minerals, whereby the rate of K^+ release not only depends on the K^+ content of the minerals but also on structural differences between K^+-bearing minerals. The K^+ status of soils mainly depends on the parent material and its degree of weathering. Whether the rate of K^+ release is adequate to meet the demand of the crop depends mainly on the kind of soil and the intensity of cropping.

Considering the mobility of soil K^+, different K^+ fractions can be distinguished: K^+ present in soil solution, K^+ adsorbed in exchangeable form to soil colloids, interlayer K^+ between the tetrahedral layers of minerals such as micas, and K^+ as a structural element of soil minerals. K^+ in soil solution and adsorbed K^+ can be easily adsorbed by plant roots (Figure 4).

K Fertilizer Recommendation

In agricultural practice it is important to know for a given location whether K^+ fertilizer application can increase crop yield. The main factors which have an impact on the crop response to K^+ fertilization are plant-available soil K^+, soil moisture, and the crop species or even sometimes different cultivars of the same species. In this respect differences in root metabolism and rooting pattern are of particular interest. Because of the longer root system of monocots like grass they profit better from soil K^+ than dicots.

High K^+ responses are likely to occur if the top soil is rather dry during the early growing stage of the crop. Because it is difficult to provide general K^+ fertilizer recommendations that are applicable to different crops, soils, and climates, the basis for recommendations should be made using soil analysis as well as K^+ fertilizer experiments for representative crops and soils. If the level of plant-available soil K^+ declines below the optimum, K^+ fertilizer must be applied. However, when K^+ fertilizers are added to soils, a portion remains in soil solution, while another portion will be held in exchangeable form and yet another portion may become unavailable, especially in soils with a high K^+ fixation capacity. Under these conditions K^+ application rates may be extremely

Table 3 Potassium (K) fertilizers

Fertilizer	*Chemical composition*	*K concentration (%)*
Muriate of potash	KCl	50, 41, 33
Potassium sulfate	K_2SO_4	43
Potassium nitrate	KNO_3	37
Sulfate potash magnesia	$K_2SO_4 \cdot MgSO_4$	18
Kainite	$KCl + NaCl + MgSO_4$	10

Adapted from Mengel K and Kirkby EA (2001) *Principles of Plant Nutrition.* Dordrecht: Kluwer Academic, with permission.

high. If the K^+ level of the soil is higher than the optimum, lower K^+ fertilizer rates are recommended or K^+ application may even be omitted. In the case of optimum K^+ conditions the amount of K^+ applied should maintain the K^+ status. This quantity can be calculated from the K^+ removal of the crop.

K Fertilizers

The usual potassium fertilizers are two main types in which K^+ is combined with either chloride (muriate of potash) or sulfate (sulfate of potash) (Table 3). Potassium chloride (KCl) is available in three different grades: 50% K, 41% K, and 33% K. The two latter forms contain substantial amounts of sodium chloride (NaCl) and are recommended as K^+ fertilizers for natrophilic crops.

Potassium sulfate (K_2SO_4) is usually manufactured by reacting the chloride with sulfuric acid. Normally K_2SO_4 contains 43% K. The fact that the use of modern highly concentrated fertilizers leads to unbalanced plant nutrition suggests that sulfate-based fertilizers, which have up to now only been used for certain special crops, may have a wider application.

Sulfate potash magnesia ($K_2SO_4 \cdot MgSO_4$) is essentially a mixture of potassium sulfate and kieserite with 18% K and 11% Mg. This is a useful fertilizer to apply to nonchloride crops when there is a need for Mg.

Potassium nitrate (KNO_3) is mainly used in intensive horticulture under glass. Low-analysis chloride-based fertilizers (kainite) are crude or partly processed mineral salts. Chemically, they are mixtures of KCl and NaCl. Some also contain kieserite, which may be part of the raw material or may have been mixed in.

The needs of the plant, soil conditions, and climatic factors will determine which form of K^+ fertilizer is best suited to obtain high yields and good quality in any particular case.

Complex Fertilizers

Complex fertilizers are stable granules with all the nutrients of the guaranteed formula. They contain two or three of the main plant nutrients N, P, and K. Sometimes secondary nutrients (calcium, magnesium, and sulfur) as well as micronutrients in soluble form are incorporated, depending on the respective formula or special grade.

NPK fertilizers commonly differ in their NPK ratios. An NPK compound 13-13-21, for example, means that the ratio of $N:P_2O_5:K_2O$ of these plant nutrients in the compound is 13% N, 13% P_2O_5, and 21% K_2O. Further NPK combinations are 10-15-20, 15-15-15, 12-12-17, 12-12-20. Combinations of N and P, N and K, as well as of N and K are also available. The complex fertilizer granules are free-flowing, resistant to physical damage and moisture, and easy to handle. The hardness of the granules as well as their resistance to abrasion keep the fertilizer from disintegration and dust formation even with frequent handling. Complex fertilizers take less time and labor to apply than individual straight fertilizers.

Bulk Blends

Bulk blends are an alternative to complex fertilizers. One of the major differences between bulk blends and complex fertilizers is the way in which the nutrients are combined. While bulk blends are a physical mixing or blending of separate dry ingredients with no chemical reaction between the different components, in complex fertilizers the chemical reaction during the production within the slurry guarantees that each granule contains the correct proportion of all nutrients.

Bulk blends are usually sold at a lower price than complex fertilizes. Their disadvantage is that bulk blends will segregate during transport, handling, and application into the individual components, due to differences in granule size and density.

Micronutrient Fertilizers

An inadequate supply of one of the micronutrients may impair plant growth and reduce yield and quality of the crop. The main issues with micronutrients are to ensure adequate but not excessive supply to plants. Micronutrients are often applied in combination with NPK fertilizers, as salts or chelates to the soil or dissolved in water sprayed on to the crop foliage (Table 4).

Iron

The most common sources of inorganic iron (Fe) sources are ferrous sulfate ($FeSO_4$) and ferrous carbonate ($FeCO_3$). Organic chelates include Fe salts of ethylenediaminetetraacetic acid (EDTA),

Table 4 Micronutrient fertilizers

Fertilizer	*Chemical composition*	*Element concentration (%)*
Ferrous sulfate	$FeSO_4 \cdot 7\,H_2O$	20
Ferrous carbonate	$FeCO_3$	48
Iron chelate	Fe-EDTA	5
Iron chelate	Fe-EDDHA	6
Manganese sulfate	$MnSO_4 \cdot 4\,H_2O$	30
Manganese oxide	Mn_2O_3, MnO_2	20
Manganese chelate	Mn-EDTA	12
Zinc sulfate	$ZnSO_4 \cdot 7\,H_2O$	23
Zinc carbonate	$ZnCO_3$	52
Zinc oxide	ZnO	80
Zinc chelate	Zn-EDTA	15
Copper sulfate	$CuSO_4 \cdot 5\,H_2O$	14
Copper oxychloride	$Cu_2Cl(OH)_3$	48
Copper chelate	Cu-EDTA	14
Boric acid	H_3BO_3	28
Borax	$Na_2B_4O_7 \cdot 10\,H_2O$	11
Solubor	$Na_2B_8O_{13} \cdot 5\,H_2O$	21
Sodium molybdate	$Na_2MoO_4 \cdot 2\,H_2O$	40
Ammonium molybdate	$3\,(NH_4)_2MoO_4 \cdot 4\,MoO_3 \cdot 4\,H_2O$	54

ethylenediamine di(*o*-hydroxyphenyl) acetic acid (EDDHA), and diethylenetriaminepentaacetic acid (DTPA).

Soils in which iron chlorosis is likely to occur are referred to as high-lime, alkaline, or calcareous. Under these conditions iron is generally not available for a continuous supply to the plant. As iron is fixed under high pH values, chelates have been shown to be superior sources of Fe for plants because lower rates are required than inorganic sources. Because of the high costs of chelates and the generally ineffective nature of inorganic Fe sources, foliar applications are the most effective method of correcting Fe chlorosis.

Manganese

Manganese sulfate ($MnSO_4$) is the most famous manganese (Mn) fertilizer. It is recommended for soil as well as for foliar application. However, when soils are treated with $MnSO_4$, banded placement rather than broadcasting should be carried out. Of the organic Mn fertilizers, Mn-EDTA, which is mainly used for foliar application, appears to give the best response. Water-insoluble Mn fertilizers are different Mn oxides (Mn_2O_3, MnO_2).

Zinc

There are many zinc (Zn) fertilizers available, but the most common inorganic compounds are zinc sulfate ($ZnSO_4$), which is the most used Zn fertilizer because of its high solubility, and zinc oxide (ZnO). Other inorganic Zn compounds are Zn carbonate ($ZnCO_3$) and Zn sulfide (ZnS). The main organic Zn fertilizer is Zn-EDTA. Both inorganic and organic Zn fertilizers are used for soil and foliar application.

Copper

Inorganic as well as organic copper (Cu) fertilizers are used to control Cu deficiency. Copper sulfate ($CuSO_4$) is the oldest Cu fertilizer, which may be used for soil and foliar application. However, caused by the scorching of $CuSO_4$ of the leaves, Cu oxychloride ($Cu_2Cl(OH)_3$) is recommended for foliar application. Cu-EDTA is recommended for soil and foliar application.

Boron

The most important boron (B) fertilizers are boric acid (H_3BO_3), borax ($Na_2B_4O_7 \cdot 10\,H_2O$), sodium tetraborate ($Na_2B_4O_7 \cdot 5\,H_2O$), sodium pentaborate ($Na_2B_{10}O_{16} \cdot 10\,H_2O$), and solubor ($Na_2B_8O_{13} \cdot 5\,H_2O$). Borax is the most used B fertilizer, and, like the other Na borates except solubur, which is applied both to soils and leaves, it is used for soil application. On high-B-fixing soils, H_3BO_3 is also used as a foliar spray. As compared with the other micronutrients the limited range between B deficiency and toxicity is extremely narrow.

Molybdenum

Molybdenum (Mo) is the micronutrient with the lowest demand. Mo deficiency is corrected by the application of sodium molybdate ($Na_2MoO_4 \cdot 2\,H_2O$) or ammonium molybdate ($3\,(NH_4)_2MoO_4 \cdot 4\,MoO_3 \cdot 4\,H_2O$). Besides soil application, foliar applications or treating seeds are common.

See also: **Fertigation**; **Fertility**; **Nutrient Availability**

Further Reading

Engelstad OP (1985) *Fertilizer Technology and Use*, 3rd edn. Madison, WI: Soil Science Society of America.

Finck A (1992) *Fertilizers and Fertilization* (in German). Weinheim, Germany: VCH.

IFA (1992) *World Fertilizers Use Manual*. Paris: International Fertilizer Industry Association.

Laegreid M, Bockman OC, and Kaarstad O (1999) *Agriculture, Fertilizers and the Environment*. New York: CAB.

Mengel K (1985) Dynamics and availability of major nutrients in soils. *Advances in Soil Science* 2: 65–131.

Mengel K (1997) Agronomic measures for better utilization of soil and fertilizer phosphates. *European Journal of Agronomy* 7: 221–233.

Mengel K and Kirkby EA (2001) *Principles of Plant Nutrition*. Dordrecht: Kluwer Academic.

Stevenson FJ (1982) *Nitrogen in Agricultural Soils*. Agronomy No. 22. Madison, WI: American Society of Agronomy, Crop Science of America, Soil Science Society of America.

Field Capacity *See* **Water Cycle**

FLOCCULATION AND DISPERSION

I Shainberg and G J Levy, The Volcani Center,
Bet Dagan, Israel

Introduction

The clay particles of soils are of colloid size (less than 2 μm) and usually carry a net negative charge. This charge is neutralized by exchangeable cations. The interaction between soil solution and soil clay and the phenomena of clay dispersion and flocculation depends on the type and amount of the soil clays, the nature of the exchangeable cations, and the electrolyte concentration in the soil solution. Clay dispersion (deflocculation) and flocculation are very important in determining the physical properties of soils such as hydraulic conductivity, aggregate stability, crusting, infiltration rate, and erosion.

In a stable clay suspension, dispersed particles collide frequently because of their Brownian motion, but separate again because of diffuse double-layer repulsion forces. When salt is added to the clay suspension, the particles may stick together upon collision, forming flocs that settle. The stability of the colloidal suspension is a balance between repulsive and attractive forces acting simultaneously among the suspended particles.

Shape and Charge of Soil Clays

The clay minerals of most soils are dominated by various layer silicates. Layer silicate clay minerals are classified as 1:1 where each layer consists of one tetrahedral silica sheet and one octahedral alumina sheet (e.g., kaolinite); 2:1 where each layer consists of one octahedral sheet sandwiched between two tetrahedral sheets (e.g., montmorillonite and vermiculite); or 2:1:1 where a metal hydroxide sheet is sandwiched between the 2:1 layers (chlorite). (*See* **Clay Minerals**.)

The clay minerals are characterized by isomorphic substitution of lower-valence cations in either the tetrahedral (Al replacing Si) or octahedral (Mg replacing Al) sheets or both. This excess of permanent negative charge inside the crystal is balanced by exchangeable cations on the external surface. Layer silicate clay minerals also possess variable charge located at the broken edges of clay minerals. At the edges of the 2:1 clay minerals, hydroxyl Al or hydroxyl Si is present. Negative charge may develop on the edges of the octahedral (Al—OH) layer at high pH and positive charge at low pH.

The colloidal activities of the clay minerals depend on their specific surface area (meters squared per gram) and the charge density expressed by their cation exchange capacity (CEC, in centimoles of cation charge per kilogram) divided by their specific surface area. Kaolinite is a widespread, 1:1 clay mineral with low CEC (1–10 $\text{cmol}_c\,\text{kg}^{-1}$), low surface area (20–50 $\text{m}^2\,\text{g}^{-1}$), and low colloidal activity. Conversely, montmorillonite, the dominant clay in soils from semiarid regions, belongs to the 2:1 clay mineral category, has a high CEC (80–120 $\text{cmol}_c\,\text{kg}^{-1}$), high surface area (800 $\text{m}^2\,\text{g}^{-1}$) and high colloidal activity. Vermiculite with high CEC (120–150 $\text{cmol}_c\,\text{kg}^{-1}$) and intermediate surface area and illite with intermediate CEC (20–40 $\text{cmol}_c\,\text{kg}^{-1}$) and intermediate surface area are clay minerals with low-to-intermediate colloidal activity. The mineralogy, charge characteristics, and CEC of several clay minerals are presented in Table 1.

Table 1 Charge characteristics and cation exchange capacities of clay minerals

	Charge per unit half-cell		*Cation exchange capacity*
Solid	*Tetrahedral*	*Octahedral*	*($\text{cmol}_c\,\text{kg}^{-1}$)*
Kaolinite	0	0	1–10
Smectite			80–120
Montmorillonite	0	−0.33	
Vermiculite	−0.85	0.23	120–150
Mica			20–40
Muscovite	−0.89	−0.05	
Chlorite			10–40

Reproduced with permission from Goldberg S, Lebron I, and Suarez DL (1999) Soil colloidal behavior. In: Sumner ME (ed.) *Handbook of Soil Science*, pp. B195–B240. Boca Raton, FL: CRC Press.

Forces Between Clay Platelets

Repulsion Forces

The electrical double layer According to the Gouy–Chapman theory, the diffuse double layer consists of the lattice charge and the compensating counterions. The counterions are subject to two opposing tendencies: the cations are attracted electrostatically to the negatively charged clay surface; and the cations tend to diffuse from the surface of the particle, where their concentration is high, into the bulk of the solution, where their concentration is lower. The two opposing tendencies result in decreasing counterion concentration from the clay surface to the bulk solution (Figure 1). Divalent cations are attracted to the surface with a force twice as large as that of monovalent cations. Thus, the diffuse double layer in the divalent ion system is more compressed toward the surface. With an increase in the electrolyte concentration in the bulk solution, the tendency of the counterions to diffuse away from the surface is diminished and the diffuse double layer compresses toward the surface. The distribution of mono- and divalent cations in the diffuse double layer at high and low electrolyte concentration is also presented in Figure 1.

In Gouy–Chapman's diffuse double-layer theory, the counterions are considered as point charges without dimensions. As long as the ions are far from the clay surface, their concentration is low and their size has little influence on the statistical calculations of the diffuse double layer. However, in the vicinity of the clay surface where the concentration of the cations is high, the distance of closest approach to the surface is important, and the size of the ions and the energy of hydration of the cations must be considered to give an accurate description of the ions' distribution near the clay surface. Consequently, the double layer is divided into the Stern layer, i.e., the first layer of the counterions immediately next to the clay surface, and the Gouy layer, which is the diffuse layer. Adsorbed cations are classified as being in the diffuse or Stern layer depending on their dimension and hydration energy. Calculations using the amended theory predict differences in the adsorption characteristics and flocculation values of Li^+, Na^+, and K^+ montmorillonite in general agreement with experimental observations.

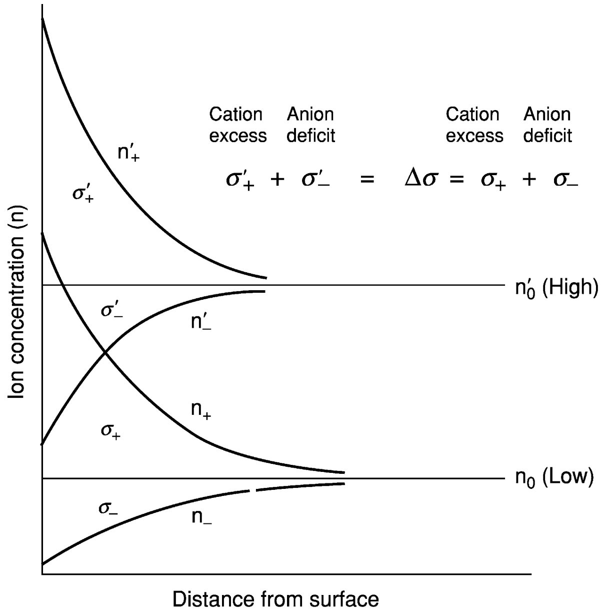


Figure 1 Distribution of counterions (n'_+ and n_+) and co-ions (n'_- and n_-) in the electrical double layer for a permanent charge surface (σ) at two electrolyte concentrations. (After Van Olphen H (1977) *An Introduction to Clay Colloid Chemistry*, 2nd edn. New York, NY: John Wiley.)

When two clay platelets approach each other, their diffuse counterion atmospheres overlap. Work must be performed to overcome the electrical repulsion forces between the two positively charged ionic atmospheres. The electrical double-layer repulsion force, also called the 'swelling pressure,' can be calculated by means of the diffuse double-layer theory. The greater the compression of the ionic atmosphere toward the clay surface, the smaller the overlap of atmospheres for a given distance between particles. Consequently, the repulsion forces between the particles decrease with increases in the salt concentration and in the valence of the adsorbed ions. Thus, high repulsion and high swelling persist in dilute solutions of sodium clay.

Hydration repulsion forces The adsorbed cations and the clay surfaces must lose some of their hydration water, because the clay particles approach each other very closely in the process of flocculation. The work that is needed to remove these water molecules manifests itself as the repulsion energy. Relative to the double-layer repulsion forces, the range of these forces is short. These forces partially negate van der Waals forces of attraction, resulting in particles not being trapped in a true primary minimum. This makes it easier for particles to disperse with little input of energy as the electrolyte concentration is reduced.

Flocculation and Attraction Forces

A stable clay suspension flocculates upon the addition of salt. Flocculation takes place only when a threshold electrolyte concentration, the 'flocculation value' (FV), is applied. A suspension of clay platelets flocculates in three possible modes of particle association (Figure 2): (1) association between flat oxygen planes

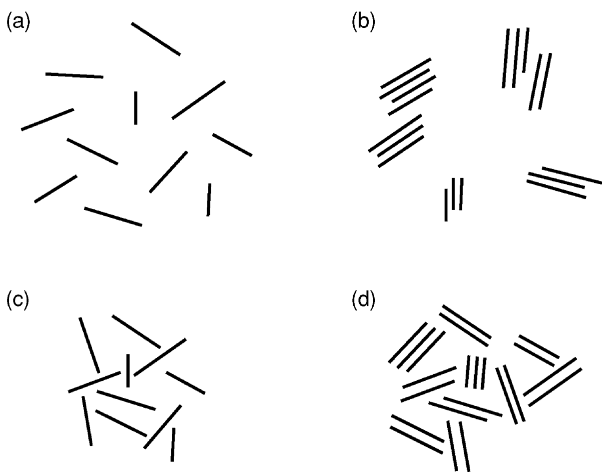


Figure 2 Schematic representation of modes of particle flocculation: (a) dispersed; (b) face-to-face flocculation; (c) edge-to-face flocculated; (d) face-to-face and edge-to-face flocculated. (After Van Olphen H (1977) *An Introduction to Clay Colloid Chemistry*, 2nd edn. New York, NY: John Wiley.)

of two parallel platelets (face to face, FF); (2) association between edge surface of neighboring particles (edge to edge, EE); and (3) association between edge and oxygen planar surface (edge to surface, EF). In the presence of low concentrations of NaCl, the diffuse double-layer repulsion prevents particle association and a stable suspension is obtained. As the concentration of NaCl increases, double layers at soil surfaces become compressed and, at the FV, attraction forces may prevail and EF, EE, and FF association may occur. Similarly, when polyvalent cations are adsorbed (e.g., Ca and Mg) on the clay surface, the diffuse double layer is compressed and attraction forces may prevail. Although the three types of association are really three modes of flocculation, only the EE and the EF types of association were discussed in the past and were labeled clay flocculation with high gel volume. The thicker particles, which result from simple FF association, were not called 'flocs,' because they were condensed with low gel volume. Special cases of FF association, which lead to high-volume flocs, have now been described (Figure 3) and it has been concluded that FF association may also lead to high-volume flocculation.

Van der Waals force of attraction The van der Waals attractive force between colloidal particles arises from the fluctuating dipole of one atom that polarizes another, and the two atoms attract each other. This attraction between atom pairs is additive and therefore the energy of interaction between particles decreases much more slowly with distance than that between individual atoms. The van der Waals attraction force leads to FF clay association.

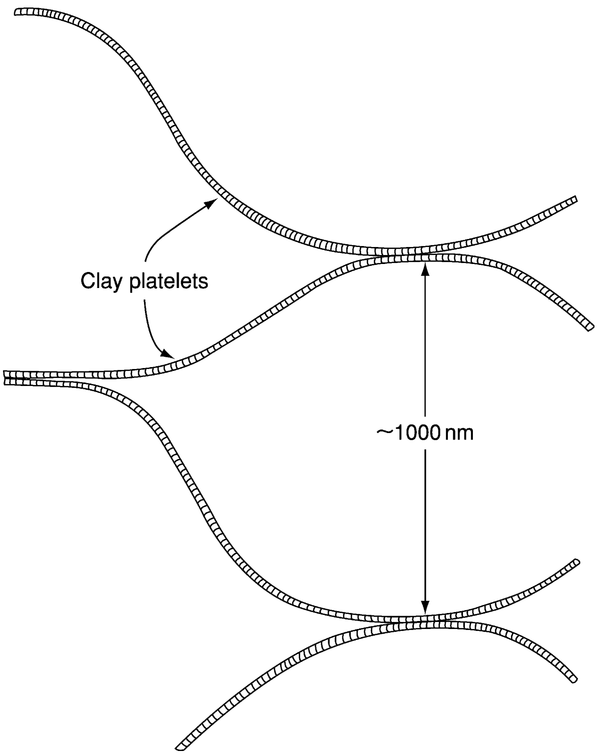


Figure 3 Mode of platelet association (face-to-face) in sodium montmorillonite suspension at moderate electrolyte concentration (e.g., 44 mmol dm^{-3} NaCl concentration). (Reproduced with permission from Keren R, Shainberg L, and Klein E (1988) Settling and flocculation value of Na-montmorillonite particles in aqueous media. *Soil Science Society of America Journal* 52: 76–80.)

Electrostatic EF attraction forces The flat surfaces are not the only surfaces of the plate-like clay particles; they also expose an edge surface area. At the edges, the octahedral alumina or iron sheets are disrupted and primary bonds are broken. This situation is analogous to that on the surface of alumina or iron particles in alumina and iron sols. On such surfaces an electric double layer is created by the adsorption of potential-determining ions. The hydroxyl-aluminum edge carries a positive charge in acid solution and a negative charge in high-pH solution. The zero point of charge does not necessarily occur at neutral pH. The charge becomes more positive with decreasing pH, and its sign may be reversed with increasing pH. When the charge on the edges is positive and the repulsive diffuse double-layer forces are diminished by a high concentration of electrolytes, a positive-edge to negative-face surface predominates and EF flocculation occurs.

Clay charge density and electrostatic attraction forces The colloid properties of clays depend to a large extent on the density of charge on the clay surface. This parameter can be derived from the exchange capacity per gram and the surface area per gram (Table 1). Sodium montmorillonite in dilute suspensions exists as single platelets. Calcium montmorillonite, even in distilled water, exists in packets (tactoids or quasicrystals), each consisting of several (4–9) clay platelets with a film of water 0.45 nm on each internal surface and c-spacing of less than 1.9 nm. The swelling pressure of calcium montmorillonite can be predicted from the diffuse double-layer theory if one assumes that the number of platelets in a tactoid is 5 and that only the Ca ions on the external surfaces of the tactoids form a diffuse double layer. Sodium, magnesium, and calcium vermiculite do not swell beyond 1.5 nm (two molecular layers of H_2O between the platelets), even in distilled water. The illites exist in quasicrystals with c-spacing of less than 1 nm regardless of the adsorbed cations. It has been suggested that electrostatic attraction forces, which increase with the increase in surface charge density and with increase in cation valence, prevent clay swelling in clays with high surface charge density.

The charge density at clay surfaces is an average value, and heterogeneity of the charge density due to isomorphous substitution has been determined for montmorillonites ranging between 0.7×10^{-3} and $1.5 \times 10^{-3}\,\mathrm{mmol_c\,m^{-2}}$. Wyoming bentonite is the most heterogeneous among six various clay types. It is possible, therefore, that at specific sites on the planar surfaces the charge density is very low or even zero (the surface area of these particular sites may not be detected). Thus, in the extreme situation where the charge density is very low (or even zero, such as in pyrophyllite), there are no repulsion forces (both double layer and hydration), and the van der Waals attraction forces exceed the repulsion forces and a FF association may occur (Figure 3). Similarly, in sites where the charge density is high (such as in vermiculite), the electrostatic attraction forces between the cations and the clay platelets is very high, swelling decreases, and a FF association may occur (Figure 3). Since the montmorillonite platelets are long and flexible and have the capability to bend, the platelets may form cohesive junctions at several locations of low or high surface charge densities to form a network structure, as illustrated in Figure 3. This results in a high gel volume, as demonstrated for Na-montmorrilonite at pH 9.8 (Figure 4).

Spontaneous and Mechanical Dispersion

Spontaneous dispersion often takes place without energy inputs when sodic clay is placed in water of very low electrolyte concentration. The proportion of clay particles separated in this way increases with an increase in the sodium adsorption ratio (SAR) of the soil solution and with a decrease in the electrolyte concentration.

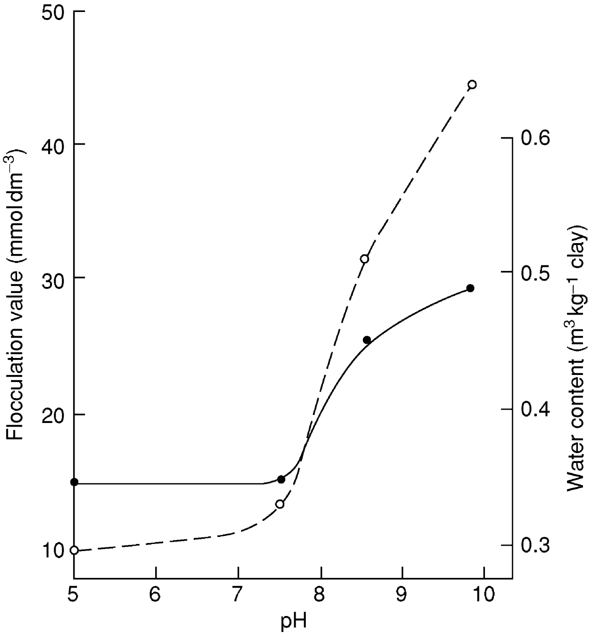


Figure 4 The flocculation value (dashed line) and gel volume (solid line) of sodium montmorillonite in NaCl solution as affected by suspension pH (the clay concentration was 0.1% and the apparent size of the platelets was less than 2 μm). (Reproduced with permission from Keren R, Shainberg L, and Klein E (1988) Settling and flocculation value of Na-montmorillonite particles in aqueous media. *Soil Science Society of America Journal* 52: 76–80.)

Hydrated clay particles in equilibrium with low-SAR solution, which have undergone limited separation, can be pushed further apart by applying external mechanical pressure such as stirring or raindrop impact. When the SAR of the clay solution is low to moderate and the electrolyte concentration is high enough to prevent spontaneous dispersion, the introduction of a mechanical force such as mixing the suspension may cause clay dispersion. The clay at the soil surface is especially susceptible to dispersion because of the mechanical energy applied by raindrops. This is one of the reasons why the soil surface is more susceptible to clay dispersion and aggregate disintegration by low sodic conditions, compared with the soil's profile.

Flocculation and Dispersion in Na/Ca Clays

The flocculation value of sodium and calcium montmorillonites Adsorbed Na ions form a diffuse double layer at the clay surface, and, because of the high swelling between the platelets, single platelets tend to persist in dilute solution. When salt is added to the clay suspension, the particles stick together upon collision, forming flocs that settle. The suspension is then separated into the bottom sediment and a

Table 2 Selected critical flocculation concentrations for some clay minerals

Mineral	*CFC (mol m^{-3})*	*Background electrolyte*	*pH*	*Suspension strength (g kg^{-1})*
Georgia kaolinite	5	NaCl	7	0.6
	245	$NaHCO_3$	8.3	0.6
	0.4	$CaCl_2$	7	0.6
Montmorillonite	12	NaCl	?	1
	0.13	$CaCl_2$	?	1
Vermiculite	38	NaCl	7	0.6
	58	$NaHCO_3$	8.3	0.6
	0.8	$CaCl_2$	7	0.6
Grundy illite	7.24	NaCl	6	1
	0.2	$CaCl_2$	6	1
Illite 36	9	NaCl	7	0.6
	185	$NaHCO_3$	8.3	0.6
	0.13	$CaCl_2$	7	0.6

CFC, critical flocculation concentration.

Based on data from Goldberg S, Lebron I, and Suarez DL (1999) Soil colloidal behavior. In: Sumner ME (ed.) Handbook of Soil Science, pp. B195–B240. Boca Raton., FL: CRC Press; Oster JD, Shainberg I, and Wood JD (1980) Flocculation value and gel structure of Na/Ca montmorillonite and illite suspensions. *Soil Science Society of America Journal* 44: 955–959.

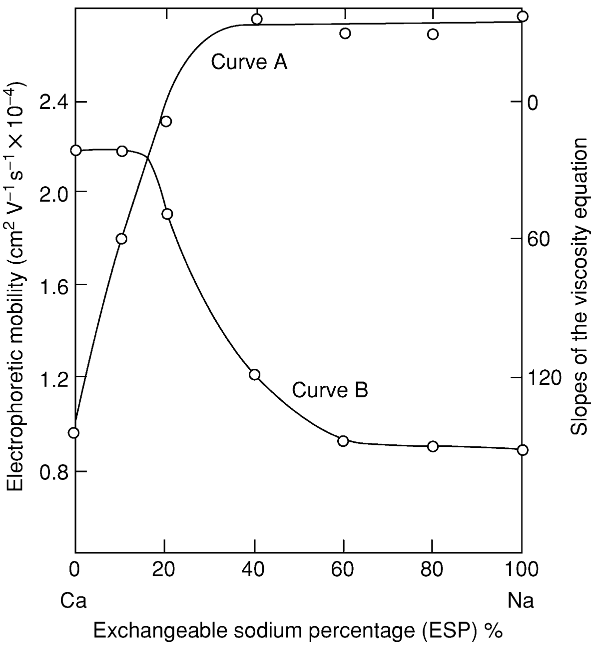


Figure 5 Dependence of electrophoretic mobility (curve A) and the relative size (curve B) of montmorillonite particles on the exchangeable sodium percentage (ESP). The relative size is expressed in units of the slope in Einstein's equation for the viscosity of suspension. (Reproduced with permission from Shainberg I and Letey J (1984) Response of soils to sodic and saline conditions. *Hilgardia* 52: 1–57.)

particle-free supernatant liquid. The FVs of clays depend on counterion valency. The FVs of sodium and calcium montmorillonite are 12 mmol$_c$ l^{-1} NaCl and 0.25 mmol$_c$ l^{-1} $CaCl_2$, respectively (Table 2). The sodium montmorillonite gel has approximately 330 g of water per gram of clay, which corresponds to a film thickness of 440 nm per surface (Figure 4). This thickness exceeds the range of diffuse double-layer forces and/or the van der Waals forces. The open structure typical of sodium montmorillonite gels suggests that EF attraction forces or the flexible platelets mechanisms (Figure 3) are operating.

Effect of pH on the FV of sodium montmorillonite

The FV of sodium montmorillonite by NaCl is pH-dependent and its value is 12.6 mmol$_c$ l^{-1} at pH 7, and 10, 13, 31, and 44 mmol l^{-1} of NaCl at pH 5, 7.5, 8.5, and 9.8, respectively (Figure 4). The final gel volume of sodium montmorillonite samples at the above pH values is high (350–490 g of water per gram of clay) (Figure 4). The high gel volume indicates that a three-dimensional 'card house' structure is formed, even at pH 9.8, in spite of the fact that at pH 9.8 a negative charge predominates at the broken edges. The high gel volume at pH 9.8 suggests that an open structure with FF association between the platelets (at sites of low and high charge density) predominates rather than the EF card house structure.

Dispersion and flocculation in Na/Ca clay systems

Whereas single platelets tend to persist in a dilute solution of sodium montmorillonite, calcium montmorillonite platelets aggregate into tactoids or quasi-crystals. Each tactoid consists of several (4–9) clay platelets in parallel array, with interplatelet distances of 0.9 nm. A diffuse double layer exists only on the external surfaces of the tactoids. In a suspension containing a mixture of Na and Ca ions, 'demixing' of the cations occurs so that some interlayer spaces contain mainly Na ions and others mainly Ca ions. An indication of the location of the adsorbed ions in montmorillonite clay saturated with a mixture of mono- and divalent cations is obtained by measurements of the electrophoretic mobility of clay particles in suspension (Figure 5). A slight addition of exchangeable Na (e.g., 10%) to Ca-saturated clay has a considerable effect on the electrophoretic mobility of the clay (Figure 5). When the exchangeable sodium percentage (ESP) in a clay suspension is less than 40%, the Na ions concentrate on the external surfaces of the tactoids and the Ca ions concentrate on the internal surfaces.

Results of studies of the FVs of montmorillonite and illite suspensions saturated with mixtures of Na and Ca ions in the exchange phase (Figure 6) indicate that a small addition of exchangeable Na to Ca-saturated montmorillonite can have a considerable effect on the FV. The demixing model, which suggests that Na ions concentrate on the external surfaces of the tactoids

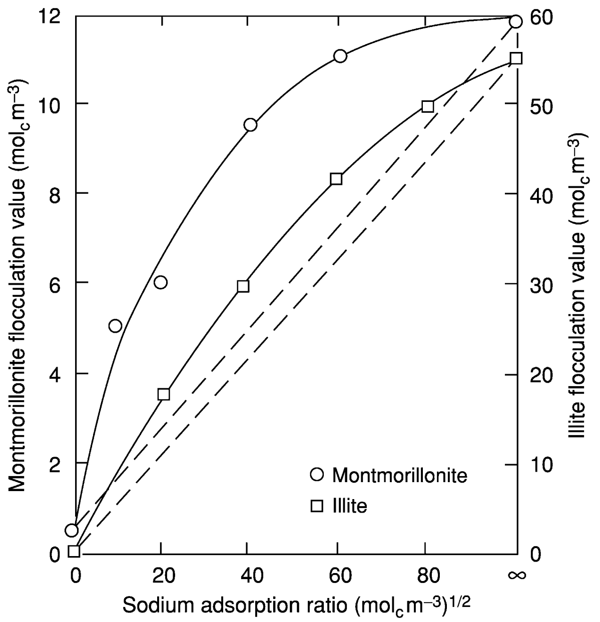


Figure 6 Flocculation value for Wyoming montmorillonite (API 25) and Fithian illite (API 35) as a function of the sodium adsorption ratio. (Reproduced with permission from Shainberg I and Letey J (1984) Response of soils to sodic and saline conditions. *Hilgardia* 52: 1–57.)

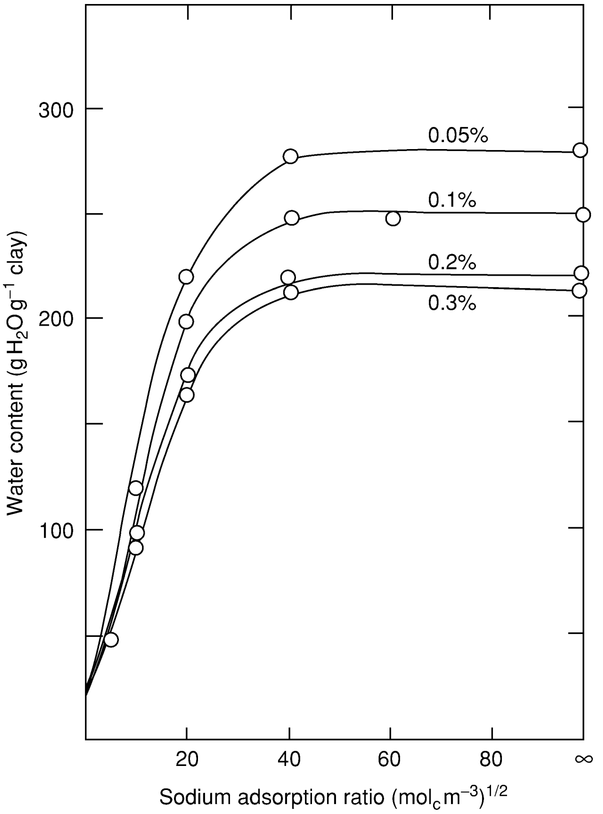


Figure 7 Water content of Wyoming montmorillonite gel as a function of Na adsorption ratio and suspension clay content. (Reproduced with permission from Oster JD, Shainberg I, and Wood JD (1980) Flocculation value and gel structure of Na/Ca montmorillonite and illite suspensions. *Soil Science Society of America Journal* 44: 955–959.)

and increase the flocculation values of the clay considerably, was applied. When the ESP of the clay suspension reaches 50, the FV of the Na/Ca mixture reaches the FV of pure sodium montmorillonite (Figure 5). Also the gel volume increases with an increase in ESP until it reaches the gel volume of pure sodium montmorillonite at ESP 50 (Figure 7). In clay suspensions with ESP of more than 50, disintegration of the tactoids is complete and single platelets predominate.

The effect of peptizing agents Sodium polymetaphosphate ($(NaPO_3)_{13}$) is a very effective peptizing agent. Organic anions may also act as peptizing agents. The FV of the peptizer-treated clay suspensions increases rapidly with increasing peptizer concentration. The anions are probably chemisorbed at the edges of the clay particles by reacting with the exposed Al. Thus, the charge of the edges becomes negative, which prevents the EF bond from taking place.

If the EF bond is eliminated by the adsorption of the polymetaphosphate, then one might expect that the open, card-house structure of the gel is also eliminated. However, an open structure with high gel volume has been obtained for the sodium gel (Table 3). This also indicates that a FF bond may lead to open gel structure.

Soil clay systems Flocculation and dispersion behavior of soil clays differs significantly from that of pure clay systems, possibly because soil clays usually occur as mixtures and because of their association with other minerals and organic matter present in the soil. A few studies demonstrate that the FV values of soil clays is two- to tenfold higher than for pure clay systems (Figure 8). Conversely, in the presence of hydrous oxides or sparingly soluble minerals such as $CaCO_3$, clay dispersivity is reported to be less severe. Hence, extrapolation from pure clay systems is of limited practical value.

Dispersion and the Hydraulic Properties of Soils

Figure 9 shows the importance of clay dispersion and movement in determining the hydraulic conductivity of soils. In this experiment a noncalcareous loam with 20% clay (montmorillonite) and a calcareous clay soil (50%, montmorillonitic) were equilibrated with a solution of 20 mmol$_c$ l^{-1} of SAR 10 to give soils with ESP 10. Following equilibration, the soils were leached with distilled water. The loam was completely sealed and the clay soil maintained a relative hydraulic conductivity of about 20% (Figure 9). This observation contradicts the generalization that soils having a higher content of expansible clays are more susceptible to sodic conditions because of more swelling. In the calcareous clay soil, salt concentration in the soil solution was maintained at 4 mmol$_c$ l^{-1} by dissolution of carbonates. At ESP 10, this concentration prevents clay dispersion, and the hydraulic conductivity decrease is due only to swelling. Conversely, in the noncalcareous loam, soil solution concentration was below 1 mmol$_c$ l^{-1}, thus clay dispersion has occurred and the dispersed clay has clogged the conducting pores.

The infiltration rate of the soil surface, exposed to rain, is more susceptible to dispersion and sodic conditions than the hydraulic conductivity of the soils

Table 3 The effect of sodium polymetaphosphate ($(NaPO_3)_{13} = 102 \times 13$) on the flocculation value and the water content of the montmorillonite gel prepared from suspensions that contained 0.1% and 0.2% clays

Peptizer concentration		NaCl flocculation value (mol$_c$ m^{-3})	Water content of gel (g H$_2$O g^{-1} clay)			
			With peptizer		Without peptizer	
ppm	mol$_c$ m^{-3}		0.1%	0.2%	0.1%	0.2%
0	0	12	240	220	240	220
1	0.01	20	260	235	200	190
10	0.1	80	170	180	150	140
100	1.0	120	160	165	130	130

Reproduced with permission from Oster JD, Shainberg I, and Wood JD (1980) Flocculation value and gel structure of Na/Ca montmorillonite and illite suspensions. *Soil Science Society of America Journal* 44: 955–959.

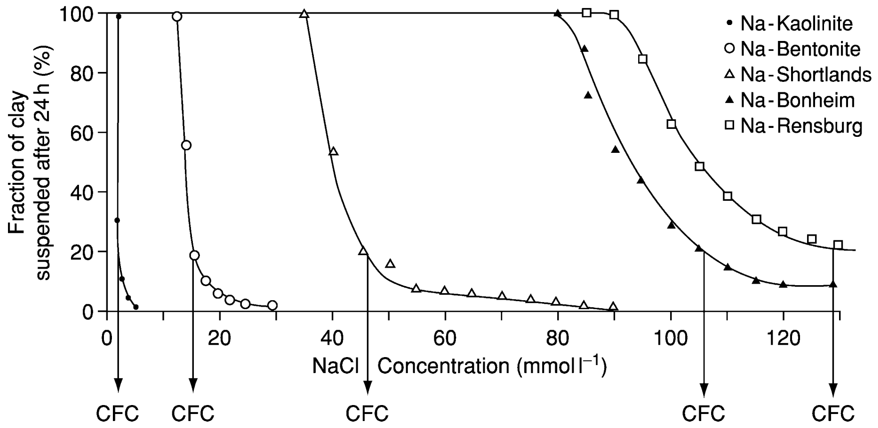


Figure 8 Flocculation behavior of Na clays suspended in NaCl solution. (Reproduced with permission from Frenkel H, Fey MV, and Levy GJ (1992) Organic and inorganic anion effects on reference and soil clay critical flocculation concentration (CFC). *Soil Science Society of America Journal* 56: 1762–1765.)

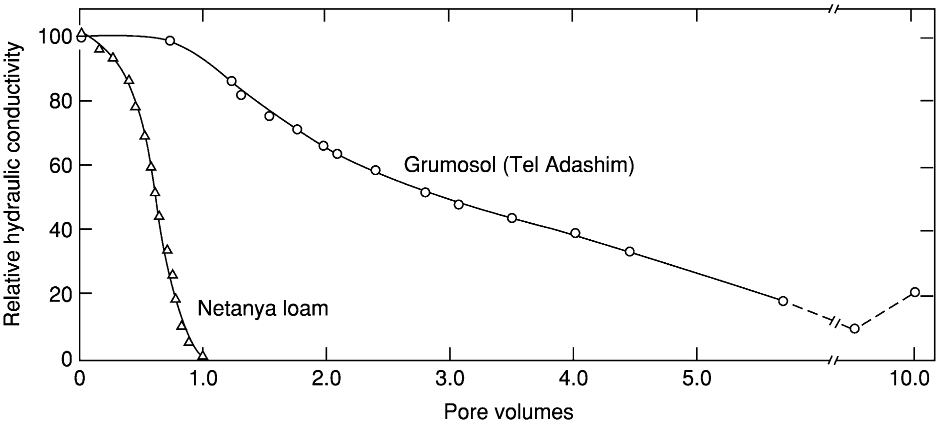


Figure 9 Hydraulic conductivity of two montmorillonite soils having an ESP of 10 and leached with distilled water. (Shainberg and Oster, 1978.) (Reproduced with permission from Shainberg I and Letey J (1984) Response of soils to sodic and saline conditions. *Hilgardia* 52: 1–57.)

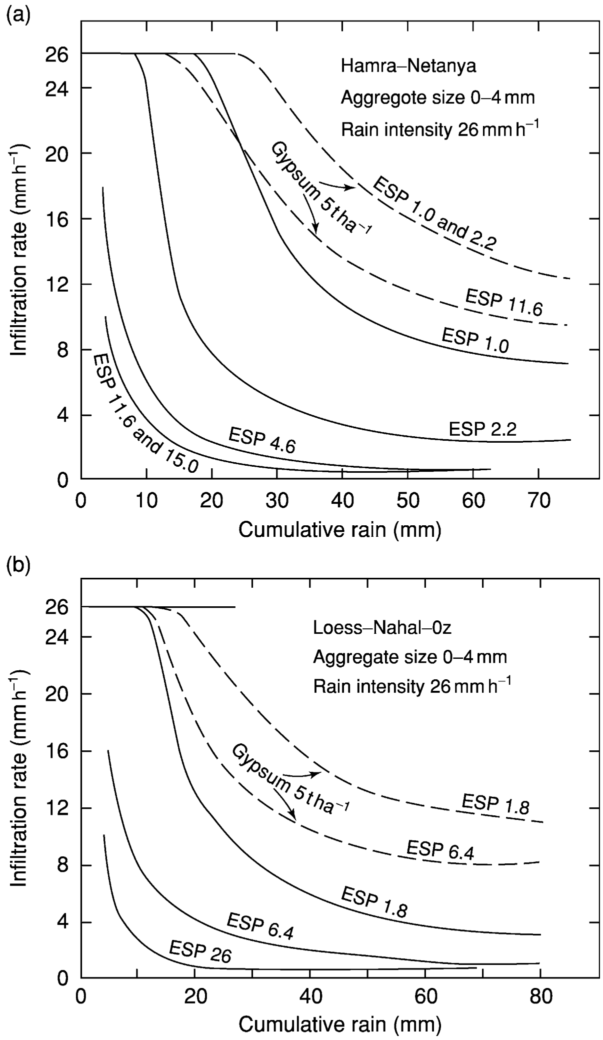


Figure 10 Effect of soil ESP (and phosphogypsum) on the infiltration rate of the (a) Hamra (Netanya) and (b) loess soil. (Adapted from Shainberg I and Lety J (1984) Response of soils to sodic and saline conditions. *Hilgardia* 52: 1–57, with permission.)

(Figure 10). Furthermore, the infiltration rates of calcareous and noncalcareous soils are equally sensitive to low levels of exchangeable sodium, whereas the hydraulic conductivity of the calcareous soils is less sensitive to low ESP than that of the noncalcareous soils. Both phenomena are explained by the clay-dispersion mechanism. At the soil surface, the concentration of the soil solution is determined solely by rain water and the low concentration of the salt enhances clay dispersion. Also, the impact of raindrops enhances clay dispersion. Both mechanisms enhance clay dispersion, aggregate disintegration, low infiltration rate, and high runoff.

Summary

The dispersion and flocculation of clays depend on the mineralogy of the clay and are affected by the composition of exchangeable cations, by the concentration of salt, by the pH, by the presence of peptizing agents, and by the application of a stirring mechanism. In a mixed Na/Ca suspension, the flocculation value of the clay increases sharply with the introduction of a small amount of sodium and with an increase in pH. Dispersion, movement, and deposition of soil colloids alter the geometry of soil pores and affect the hydraulic properties of soils. Thus, introducing a small amount of Na to soils exposed to leaching with rain water may cause clay dispersion and a deterioration of soil physical properties.

See also: **Clay Minerals**

Further Reading

Frenkel H, Fey MV, and Levy GJ (1992) Organic and inorganic anion effects on reference and soil clay critical flocculation concentration. *Soil Science Society of America Journal* 56: 1762–1766.

Goldberg S, Lebron I, and Suarez DL (1999) Soil colloidal behavior. In: Sumner ME (ed.) *Handbook of Soil Science*, pp. B195–B240. Boca Raton, FL: CRC Press.

Keren R, Shainberg I, and Klein E (1988) Settling and flocculation value of Na-montmorillonite particles in aqueous media. *Soil Science Society of America Journal* 52: 76–80.

Olson CG, Thompson ML, and Wilson MA (1999) Phyllosilicates. In: Sumner ME (ed.) *Handbook of Soil Science*, pp. F77–F123.

Oster JD, Shainberg I, and Wood JD (1980) Flocculation value and gel structure of Na/Ca montmorillonite and illite suspensions. *Soil Science Society of America Journal* 44: 955–959.

Quirk JP (1994) Interparticle forces: a basis for the interpretation of soil physical behavior. *Advances in Agronomy* 53: 121–183.

Rengasamy P and Sumner ME (1998) Processes involved in sodic behavior. In: Sumner ME and Naidu R (eds) *Sodic Soils*. New York: Oxford University Press.

Shainberg I and Kemper WD (1996) Hydration status of absorbed ions. *Soil Science of Society of America Proceedings* 30: 707–713.

Shainberg I and Letey J (1984) Response of soils to sodic and saline conditions. *Hilgardia* 52: 1–57.

Stul MS and Mortier WJ (1974) The heterogeneity of the charge density in montmorillonite. *Clays and Clay Minerals* 22: 391–396.

Van Olphen H (1977) *An Introduction to Clay Colloid Chemistry*, 2nd edn. New York, NY: John Wiley.

FLUORESCENCE SPECTROSCOPY

N Senesi and V D'Orazio, Università di Bari,
Bari, Italy

Introduction

Classical monodimensional fluorescence spectroscopy in the emission, excitation, and synchronous-scan modes and total luminescence or tridimensional fluorescence are noninvasive, nondestructive, nonseparative, and sensitive techniques requiring minimal sample pretreatment. They have proven to be very promising and useful in the study of both organic and inorganic soil constituents. The variety of fluorescent structures indigenous in soil organic components, and especially in its humified fractions, i.e., humic substances (HS), mainly humic acids (HAs) and fulvic acids (FAs), has provided invaluable information on their structural and functional chemistry and reactivity as a function of relevant environmental factors such as pH and ionic strength. The capacity of metal ions to quench soil HS fluorescence has provided new insights into various aspects of metal–HS interaction/complexation processes. Fluorescence quenching and fluorescence polarization techniques have been applied successfully for mechanistic and quantitative studies of binding/adsorption of several organic pollutants to soil HS. Further, fluorescent organic probe molecules have proven to be very useful in studies of soil mineral–solution interfaces in various conditions. Recent advances in commercially available fluorescence instrumentation and in fluorescent derivatizing agents have opened up further possibilities which were previously unattainable.

Basic Principles, Methodology, and Instrumentation

The Fluorescence Phenomenon

Every molecule possesses a series of closely spaced energy levels (Figure 1). When a quantum of light impinges on a molecule and is absorbed, the electron is promoted from the groundstate, S_0, to an excited singlet state, S_1, S_2, etc. Typical transitions of this kind are very probable in complex molecular systems containing atoms with lone pairs of electrons, such as O and N, and aromatic and/or aliphatic conjugated

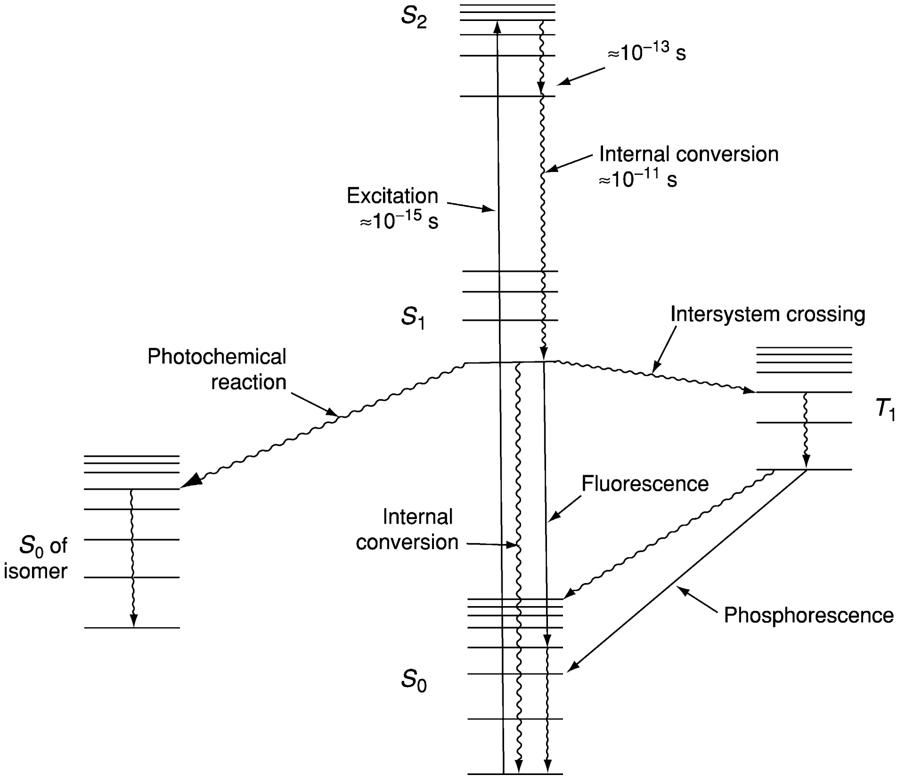


Figure 1 Schematic energy-level diagram for a hypothetical molecule.

unsaturated systems capable of a high degree of electron delocalization. Any excited state then relaxes down to the lowest vibrational level of that state, and, subsequently, energy is released by ‘internal conversion,’ dropping the electron to the lowest vibrational level of the lowest excited state, S_1. Relaxation from the S_1 state, which has a lifetime of the order of 10^{-9} s, to the groundstate S_0 can then occur by different mechanisms (Figure 1). These include: (1) the radiative photoprocess of fluorescence, which consists in the emission of photons less energetic than those absorbed as a result of the energy loss incurred during internal conversion processes; and (2) various nonradiative processes, such as: (a) internal conversion processes occurring by relaxation through internal vibrations or through collisions with solvent and/or solute molecules; (b) transfer of energy to another chemical compound resulting in a photochemical reaction or formation of an excited-state dimer (excimer) or excited-state complex (exciplex) which then emits photons at longer wavelengths than fluorescence; (c) intersystem crossing to a triplet state, T_1, that has a lifetime much longer (often $>10^{-5}$ s) than its S_1 precursor, which is followed by either a delayed release of energy known as the phenomenon of phosphorescence, or by relaxation through internal conversion processes.

Instrumentation and Experimental Procedures

Generally, minimal sample preparation is required for fluorescence measurements. Solutions and dilute colloidal suspensions (<1 g solid l^{-1}) can be placed directly into quartz cells, whereas the use of dense colloidal suspensions should be avoided because these result in increases in light scattering and light absorption, and a subsequent decrease in signal-to-noise ratio.

The principal components of a fluorescence spectrophotometer used for steady-state measurements are shown schematically in Figure 2. Excitation energy is provided by a high-energy continuum light source, commonly a xenon discharge lamp. The desired excitation wavelength is selected by passing incident radiation through a monochromator before entering the sample cell, preferentially a quartz cell. Radiation emitted by the sample is generally monitored at an angle of 90° to the excitation radiation to minimize signal distortion due to light scattering, or using front-surface detection. Wavelength selection of the emitted radiation is obtained with another monochromator. The emitted light is then measured by the detector, generally a photomultiplier tube. The instrument is completed by a readout device, generally a recorder output or, preferentially, a microprocessor for data analysis.

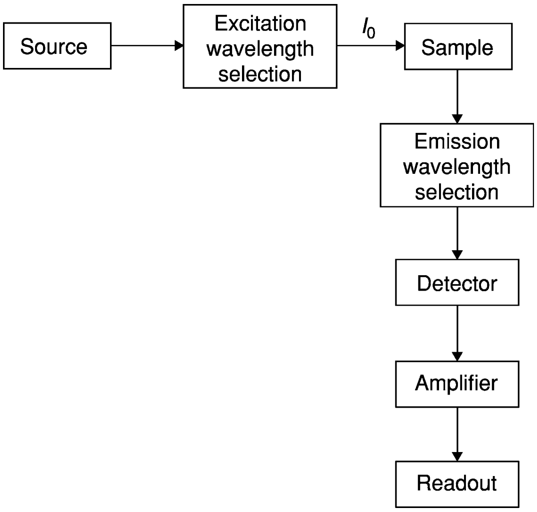


Figure 2 Schematic diagram of typical fluorescence instrumentation.

Thermostated cell holders to control sample temperature and cell stirrers can be added to most commercial spectrofluorimeters. New instrumental developments include the use of lasers as excitation sources and the use of array or image detectors. The entire instrument can be put under computer control with several benefits in precision, calculation of quantum yields, and corrections for factors affecting fluorescence, such as the inner-filter effect and photodecomposition. Due to instrumental factors such as fluctuations in source intensity, common spectrofluorimeters provide uncorrected spectra that vary from instrument to instrument; thus a correction must be made to obtain a ‘true’ spectrum in terms of quanta versus wavelength. Spectrofluorimeters that include provisions for obtaining corrected spectra have been designed.

Time-resolved fluorescence spectra or fluorescence lifetime measurements, which allow the measurement of the time a molecule resides in the excited state, involve the use of pulsed or modulated light sources and a rapid-detecting photomultiplier, a photodiode array, or a time-correlated single-photon counter. Polarized fluorescence measurements can be achieved by adapting a conventional instrument by placing appropriated oriented polarizers between the excitation monochromator and the sample, and between the sample and the emission monochromator.

Fluorescence Spectra

Conventional monodimensional fluorescence spectra can be obtained in the three modes of emission, excitation, and synchronous-scan excitation on samples dissolved or suspended in aqueous media. The emission spectrum is recorded by measuring the relative intensity of radiation emitted as a function of the wavelength at a constant excitation wavelength. The excitation spectrum is recorded by measuring the emission intensity at a fixed wavelength while varying the excitation wavelength. Synchronous-scan excitation spectra are obtained by measuring the fluorescence intensity while simultaneously scanning over both the excitation and emission wavelengths, and keeping a constant, optimized wavelength difference, $\Delta\lambda = \lambda_{\mathrm{em}} - \lambda_{\mathrm{exc}}$, between them. The synchronous-scan technique can selectively increase the intensity of specific peaks, thus improving the spectral structure resolution by decreasing the intrinsic broadness and band overlapping occurring in emission and excitation spectra of molecularly complex and heterogeneous multicomponent samples.

Total luminescence or tridimensional (3D) fluorescence spectra provide a complete representation of the fluorescence spectral features of a sample in the form of an excitation–emission matrix (EEM), in which fluorescence intensity is measured as a function of the excitation wavelength on one axis and emission wavelength on the other. To generate EEMs a series of emission spectra over a preselected wavelength range (generally chosen within 350–600 nm) are collected over a range of excitation wavelengths (generally chosen within 200–500 nm) by selecting an appropriate wavelength step size for both the excitation (e.g., from 4 to 10 nm) and the emission (generally, 0.5 or 1 nm). Fluorescence intensity data sets at each excitation wavelength are then converted into 3D spectra and EEMs using adequate macroprograms, and analyzed using dedicated computer graphic softwares.

The EEM spectra can be compared using the uncorrected matrix correlation (UMC) method. The UMC value equals 1 where the two matrices are identical and decreases to approaching zero with decreasing similarity between them. The recently developed fluorescence zone integration (FZI) method, which utilizes quantitatively all the wavelength-dependent fluorescence intensity data from an EEM spectrum, is capable of quantifying differences in EEM among various samples, and to measure the continuum of binding sites in metal and organic complexes of a sample.

Fluorescence Efficiency, Lifetime, and Intensity

Fluorescence efficiency, or quantum yield, Φ_{f}, measures the effectiveness with which the adsorbed energy is reemitted, and is defined as the ratio of the total energy emitted as fluorescence per total energy adsorbed. Absolute efficiency measurements are difficult because they require calibrated sources and detectors. However, the relative fluorescence efficiency of a sample, Φ_{s}, can be easily obtained by measuring the fluorescence intensity of a dilute solution (I_{k}) of a standard of known efficiency, Φ_{k}, such as quinine sulfate, and that of the sample (I_{s}):

$$\Phi_{\mathrm{s}} = \Phi_{\mathrm{k}}(I_{\mathrm{s}} q_{\mathrm{k}} A_{\mathrm{k}} / I_{\mathrm{k}} q_{\mathrm{s}} A_{\mathrm{s}}) \qquad [1]$$

where I values are determined by integrating the area beneath the corrected fluorescence spectrum, q is the relative photon output of the source at the excitation wavelength taken directly from the curve, and A is the absorbance.

Fluorescence lifetime, τ, refers to the probability of finding a given molecule in the excited state at time t after the excitation source is turned off, and represents the mean lifetime of the excited state. The value of τ can be calculated by measuring the fluorescence intensity I at time t, $I(t)$, according to:

$$I(t) = I_{\mathrm{M}} \exp(-t/\tau) \qquad [2]$$

where I_{M} is the maximum intensity with the excitation source on, i.e., during excitation, and t is the time elapsed after the excitation source is turned off.

Fluorescence intensity, I, depends on the molar concentration of adsorbing species in solution, C, and the fluorescence efficiency, Φ_{f}, according to:

$$I = \Phi_{\mathrm{f}} I_0 [1 - \exp(-\epsilon b C)] \qquad [3]$$

where I_0 is the intensity of incident radiation, ϵ is the molar absorptivity at the excitation wavelength, and b is the path length of the cell. For very dilute solutions, where ϵbC is sufficiently small, eqn [3] reduces to a linear relationship:

$$I = \Phi_{\mathrm{f}} I_0 \epsilon b C \qquad [4]$$

and the fluorescence intensity is essentially homogeneous throughout the sample. When ϵbC increases, the fluorescence intensity is no longer homogeneous within the cell but is increasingly localized at the front surface of the cell, thus the fraction of fluorescence measured by the common detection system is smaller (inner-cell effect). To avoid this effect, instruments are designed to look at an area in the front part of the cell rather than at the center.

Limitations and Problems

Fluctuations in intensity of the exciting light source, variations in instrumental sensitivity of the detection system, and other instrumental factors may cause distortions and variations of instrumental response. The shape of the excitation spectrum should be identical to that of the absorption spectrum of the molecule, but differences may be observed due to instrumental artifacts arising from variations in source intensity with wavelength and other causes. The shapes of emission and excitation spectra should be independent of the wavelength of the exciting radiation and the wavelength at which fluorescence is measured, but spectra are often wavelength-dependent. In practice, however, the variations expected are likely to be small when the fixed excitation or emission wavelengths used are varied within a limited range of values, i.e., between 320 and 370 nm and 520 and 570 nm, respectively. Light-scattering effects such as Rayleigh, Tyndall, and Raman effects may also influence to a varying extent the fluorescence spectra recorded.

Corrections accounting for instrumental artifacts and for scattering effects should be applied to fluorescence spectra, particularly when quantitative comparisons are made between spectra measured on a variety of spectrofluorimeters by different operators. However, comparisons of uncorrected fluorescence spectra may be allowed on a qualitative basis when they are recorded on the same instrument using the same experimental conditions.

Molecular and Environmental Factors Affecting Fluorescence

Several molecular parameters can affect fluorescence intensity and wavelength. In particular, atoms and ions of large atomic number, such as halogens and paramagnetic metals, decrease the fluorescence efficiency of a fluorescent organic compound. Increasing the extension of the π-electron system increases fluorescence and shifts the emission wavelengths toward longer values. Steric hindrance generally reduces fluorescence, whereas increasing molecular structural rigidity has the effect of increasing fluorescence. Electron-withdrawing groups, such as carbonyls and carboxyls, largely reduce fluorescence intensity, whereas the opposite is true for electron-donating substituents such as hydroxide, alkoxide, and amino groups. Further, all these groups shift fluorescence to longer wavelengths.

A number of environmental factors can also affect the fluorescence response of a sample. For example, fluorescence efficiency is decreased by temperature increase, or by the presence of a heavy atom in the solvent molecule, whereas it is increased by increasing solvent viscosity. Fluorescence wavelength is shifted to shorter or longer values by an increase in solvent polarity. Fluorescence intensity is reduced by increasing solute–solvent interaction, e.g., by hydrogen bonding. Therefore, heavy-atom solvents and hydrogen-bonding solvents should be avoided whenever possible.

Light absorption by the solvent, large concentrations of a fluorescent solute, the presence in solution of a solute that absorbs radiation together with the fluorescent molecule, or a solute that absorbs the fluorescence of the sample molecule, give rise to the so-called 'inner-filter' effect that causes quenching of fluorescence of the sample. This effect can be reduced by selecting a suitable excitation wavelength, diluting the sample, increasing solvent viscosity, viewing fluorescence closer to the front surface of the cell, and/or using the method of standard additions or a correction factor.

The pH of the medium affects fluorescence spectra of acidic or basic aromatic compounds by rapid proton transfer reactions occurring to various extents during the lifetime of the excited state. Dissolved molecular oxygen can quench fluorescence of some organic molecules by several mechanisms, especially in polar solvents. This effect can be easily eliminated by bubbling nitrogen through the solution. Metal ions, especially paramagnetic ions, are also fluorescence quenchers, even where they do not form complexes with the fluorescent molecule. Metal quenching can be reduced markedly, e.g., by diluting the sample.

Fluorescence Quenching

The emission intensity of a fluorescent species can be quenched by its association and/or adsorption to organic and inorganic substrates. The 'simple' binding ($P-S$) between a fluorescent organic pollutant, P, and a single type of site of a sorbent, S, can be represented by:

$$P+S=P-S \qquad [5]$$

with a binding constant, K_b:

$$K_b=[P-S]/[P]\,[S] \qquad [6]$$

The mass balance on P is described by:

$$[P_{tot}]=[P]+[P-S] \qquad [7]$$

where $[P_{tot}]$ is the formal or total concentration of P in the system.

Combining eqns [6] and [7] and rearranging yields:

$$[P_{tot}]/[P]=1+K_b[S] \qquad [8]$$

Assuming that the fluorescence intensity is proportional to the concentration of free P in solution, $[P]$, then:

$$I_0/I=1+K_b[S] \qquad [9]$$

where I_0/I is the ratio of fluorescence intensities of the pollutant, P, in the absence (I_0) and in the presence (I) of a quenching sorbent S. The value of K_b can thus be readily calculated from fluorescence measurements. However, in the case where S and $P-S$ include a large number of adsorbing/complexing sites and sorbates/

complexes, a rigorous interpretation of fluorescence data is much more difficult, and only the estimation of an average binding quotient is possible.

Fluorescence Polarization

The fluorescence polarization method can be applied to study the conformation of a fluorescent organic molecule and the sorption/binding of a fluorescent organic pollutant to a variety of environmental sorbents. Any process that causes a decrease in the rate of rotation of a fluorophore, i.e., a variation in its size, during the lifetime of the excited state, including change of conformation, adsorption on suspended particulate matter, or binding to dissolved materials, produces a decrease in polarization. The polarization, p, can be calculated by the fluorescence intensities measured in the planes parallel and perpendicular to the plane of polarization of the excitation radiation, $I_{\parallel}$ and $I_{\perp}$, respectively, according to:

$$p = (I_{\parallel} - I_{\perp})/(I_{\parallel} + I_{\perp}) \qquad [10]$$

By measuring the polarization values for the free/dissolved fluorophore (p_f), the bound/adsorbed fluorophore (p_b) and their mixture (p), it is possible to obtain the ratio of concentrations of bound (F_b) and free (F_f) fluorophore, according to:

$$F_b/F_f = (\Phi_f/\Phi_b)[(p - p_f)/(p_b - p)] \qquad [11]$$

where Φ_f/Φ_b is the ratio of fluorescence efficiencies for free and bound fluorophore. The values of Φ_f and p_f can be easily determined from the fluorophore solution in the absence of sorbent material, whereas the values of Φ_b and p_b can be measured in the presence of a large excess of sorbent material so that practically all the fluorophore can be considered as adsorbed. Measuring the polarization, p, as a function of binding/sorbent titrated into a fluorophore solution at constant total concentration, or vice versa, can then provide data to construct an adsorption isotherm, and possibly calculate an equilibrium constant for binding, if stoichiometry is known or can be assumed.

Fluorescence Correlation Spectroscopy

In the fluorescence correlation spectroscopy (FCS) technique a laser light (excitation at 488 or 514 nm) is focused into the sample using confocal optics, thus realizing an open, illuminated volume element that is called ‘confocal volume.’ In the absence of chemical processes, temporal variations of measured fluorescence intensity in the confocal volume can be attributed uniquely to the translational diffusion of the fluorophore. These variations can be analyzed using an autocorrelation function that includes a parameter, τ_D, which is the diffusion time characteristic of the particle under study in the confocal volume. The system can be calibrated with a rhodamine 6G (R6G) standard that has a known diffusion coefficient. The values of diffusion times for R6G and the sample in each condition are then determined from the best fit of the autocorrelation function, and finally used to calculate the diffusion coefficient of the fluorescent sample.

Soil Organic Components

Fluorescence Spectra

Although fluorescent structures constitute only a minor portion of HA and FA macromolecules, fluorescence analysis has provided unique information on their structural and functional chemistry. Fluorescence emission spectra of soil HAs and FAs generally consist of a unique broad band featuring a flat maximum at a wavelength ranging from about 500 to 520 nm for HA and from about 445 to 465 nm for FA, as a function of their origin and nature (Figure 3). The overall fluorescence intensity of FA is generally larger than that of HA.

Fluorescence excitation spectra of soil HAs and FAs generally feature a number of distinct peaks and/or shoulders whose exact position and relative intensity depend on their origin and nature. In particular, HAs generally feature two closely spaced major peaks in the long-wavelength region (around 465 and 450 nm), often accompained by a minor peak or shoulder in the intermediate-wavelength range (at 400–390 nm) (Figure 3). In contrast, FAs generally exhibit a main excitation peak in the intermediate region of the spectrum (400–390 nm), with additional minor peaks and shoulders at longer and/or shorter wavelengths (Figure 3).

Synchronous-scan spectra of soil HAs generally feature only one peak in the long-wavelength region, often accompanied by faint shoulder(s) at longer and shorter wavelengths (Figure 3). Soil FAs generally exhibit two main synchronous-scan peaks at long (450–460 nm) and intermediate (390–400 nm) wavelengths, often with some less intense peaks and shoulders at both sides (Figure 3).

Total luminescence or EEM spectra of soil HAs and FAs generally show excitation/emission fluorescence pairs in the wavelength range of 300–400/400–500 nm, which is in agreement with monodimensional spectra. Soil HAs feature fluorescence maxima at longer excitation/emission wavelengths and are less intensely fluorescent than soil FAs. Typical EEM spectra of three reference HA and FA

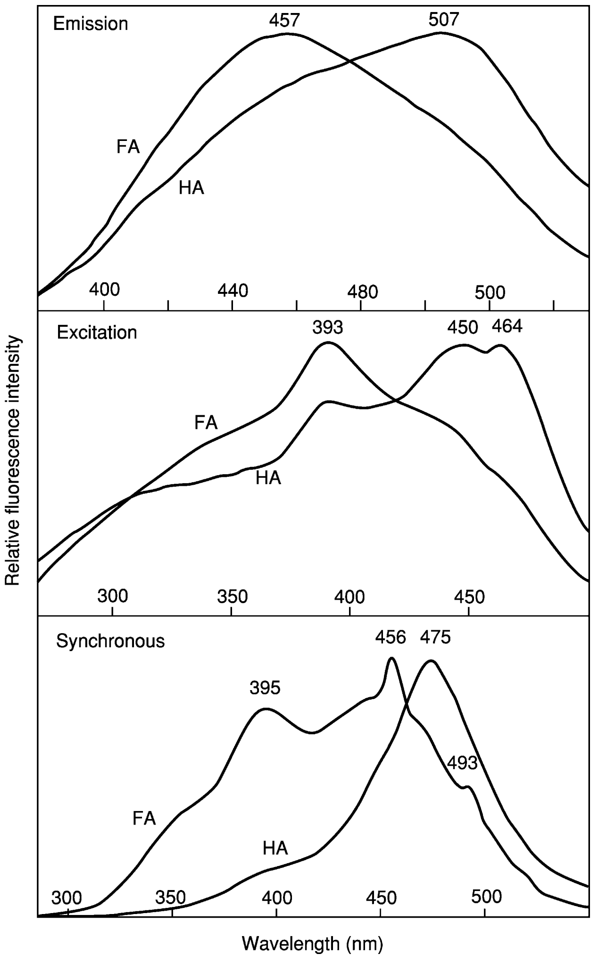


Figure 3 Fluorescence emission, excitation, and synchronous-scan spectra of the International Humic Substances Society (IHSS) reference humic acid (HA) and fulvic acid (FA) isolated from a Mollic Epipedon soil. Reproduced with permission from Senesi N, Miano TM, Provenzano MR, and Brunetti G (1991) Characterization, differentiation, and classification of humic substances by fluorescence spectroscopy. *Soil Science* 152: 259–271

samples of the International Humic Substances Society (IHSS) collection are shown in Figure 4.

The FZI method has been applied successfully to analyze EEMs of fractionated and bulk soil dissolved organic matter (DOM). Integration beneath EEM spectral regions selected on the basis of chemical criteria makes it possible to quantify DOM with similar fluorescent properties, and to interpret structural differences and the possible origin of DOM samples.

Nature of Fluorophores

Due to their molecular complexity and chemical heterogeneity, fluorescence spectra of HS probably represent the sum of the spectra of a number of different fluorophores. Although the identification of specific fluorophores in HAs and FAs is difficult, if not impossible, some hypotheses have been suggested on their possible chemical nature. In particular, the small fluorescence intensities and long wavelengths of the main fluorescence peaks of soil HAs can be attributed to large-molecular-weight (MW) components possessing linearly condensed aromatic ring systems bearing electron-withdrawing substituents, such as carbonyl and carboxyl groups, and/or to other unsaturated bond systems capable of a great degree of conjugation. Conversely, the large fluorescence intensities and short wavelengths typical of soil FAs can be ascribed to simple structural components of small MW bearing electron-donating substituents, such as hydroxyl, methoxyl, and amino groups, small degree of aromatic polycondensation, and small level of conjugated chromophores.

A number of structural components, which are known to be fluorescent as simple compounds and have been identified in HS macromolecules by

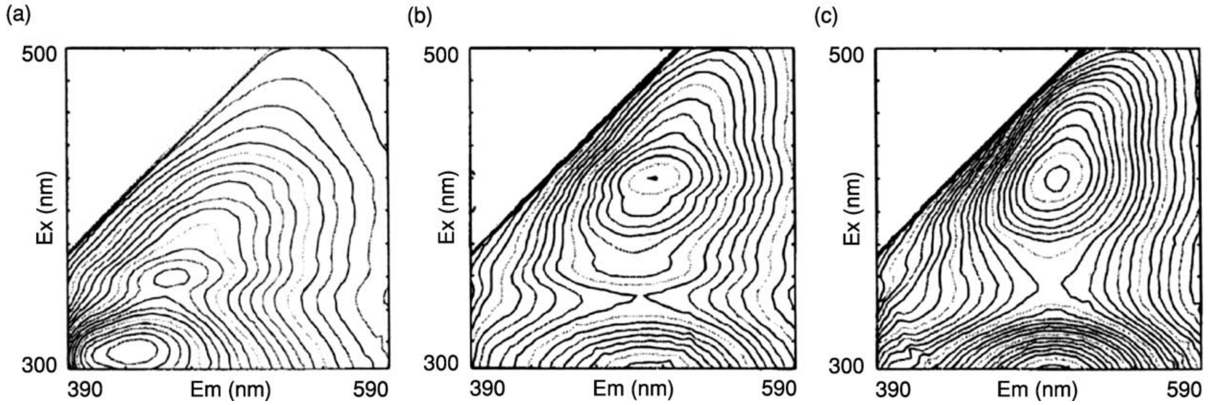


Figure 4 Absorbance-corrected excitation–emission matrix spectra of (a) a soil fulvic acid and (b and c) two soil humic acids obtained from the International Humic Substances Society (IHSS) reference collection (50 mg l^{-1}, 0.010 mol l^{-1} KCl, pH 6). Reproduced with permission from Mobed JJ, Hemmingsen SL, Autry JL, and McGown LB (1996) Fluorescence characterization of IHSS humic substances: total luminescence spectra with absorbance correction. *Environmental Science and Technology* 30: 3061–3065.

degradative methods and nuclear magnetic resonance (NMR) and infrared (IR) spectroscopies, have been suggested as potential contributors to the fluorescence of HS (Table 1). These include: (1) benzene rings bearing a hydroxyl group conjugated to a carbonyl, methylsalicylate units, and dihydroxybenzoic acid structures, such as protocatechuic, caffeic, and ferulic acids, which fluoresce at short wavelengths; (2) hydroxy- and methoxy-coumarin-like structures possibly originated from lignin, chromone, and xanthone derivatives of plant origin, and Schiff-base structures derived from polycondensation of carbonyls and amino groups, which fluoresce at intermediate wavelengths; and (3) flavonoid, naphtol and hydroxyquinoline structures that fluoresce at long wavelengths.

Molecular and Environmental Factors Affecting Fluorescence

The fluorescence intensity and spectral shape of HS can be affected by a number of molecular and environmental parameters. For example, a decrease in fluorescence intensity, shift to long wavelengths, and broadening of the emission maximum are observed with increasing MW of HS macromolecules. These effects can be related to the more extensively substituted aromatic network and greater degree of polymerization of the heavier HS fractions in comparison to the lighter fractions.

A concentration in the range 5–10 mg l^{-1} does not significantly affect the EEM spectra of HS. However, as the concentration increases from 10 to 120 mg l^{-1} at pH = 7.0 for soil FAs, fluorescence emission shifts to longer wavelengths, the overall fluorescence intensity increases and the relative intensities of synchronous-scan peaks change (Figure 5). In contrast, soil HAs fluoresce more intensely at intermediate concentrations (16–70 mg l^{-1}). These effects can be ascribed to modifications of the degree of association and/or to configuration rearrangements affecting HA and FA fluorophores.

The spectral shapes and relative fluorescence intensities can be variously affected by pH changes on dependence of the nature and origin of HS. For example, the synchronous-scan peaks of a soil FA reduce from four to two as the pH increases from 4.5 to 10 (Figure 6). Increasing the pH from 2 to 10, the fluorescence intensity increases and a red shift occurs in both the long- and short-wavelength regions of EEM spectra of some soil FAs (Figure 7). These effects can be attributed to variations in ionization of acidic groups, especially phenolic hydroxyl groups, changes in free radical concentration and/or particle association, and conformational rearrangements such as decoiling of macromolecular structures, e.g., by disruption of hydrogen bonds and consequent major exposure of functional groups to the solvent.

Ionic strength due to KCl in the range of 0–1 mol l^{-1} appears not to affect significantly the EEM spectra of HS. However, a gradual decrease in fluorescence intensity is measured for some soil HAs and FAs upon increasing ionic strength due to NaCl concentration from 0.001 m to 0.01 mol l^{-1}. This effect can be ascribed to both the gradual coiling-up of the HS macromolecular structure and the salt-depressing ionization of HS functional groups.

Increasing the temperature from 18 to 48°C causes a decrease in intensities and a shift to longer wavelengths of the major emission and excitation peaks of FAs. This effect can be ascribed to the increased deactivation of excited molecules by processes other than fluorescence, e.g., intermolecular collisions. The intensities of emission and excitation peaks of FAs tend to increase with increasing the redox potential from −1.17 to +2.5 V in solutions at small concentrations of mild reductants or oxidants.

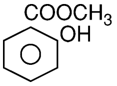
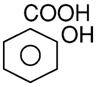
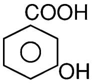
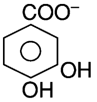
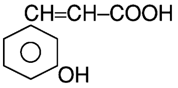
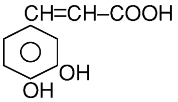
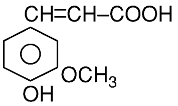
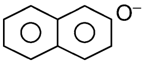
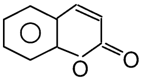
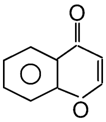
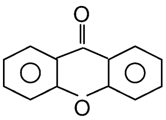
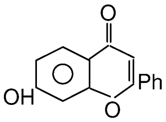
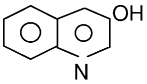
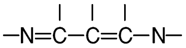
Humification Indexes and Differentiation of Humic Substances by Fluorescence Properties

Both conventional monodimensional and EEM fluorescence spectra can be used as diagnostic criteria for differentiating HA and FA from the same source and HAs and FAs of different origin, and for evaluating their humification degree. A humification index of soil DOM has been developed based on the ratio of the area under the 'upper quarter' ('H,' 435–480 nm) and the area under the 'lower quarter' ('L,' 300–445 nm) of the fluorescence emission spectrum of the sample measured at an excitation wavelength of 254 nm. Another humification index proposed for soil DOM is based on the ratio of synchronous-scan fluorescence intensity measured either at 400 and 360 nm or at 470 and 360 nm.

The humification degree of FAs of various origin has been related to their synchronous-scan spectra obtained at various pHs. Finally, a significant positive correlation has been shown to exist between the humification degree of several soil HAs, evaluated from the concentration of organic free radicals measured by electron paramagnetic resonance (EPR) spectroscopy, and total fluorescence, calculated as the area under the emission spectrum obtained at an excitation wavelength of 400 nm.

However, the classification and humification indexes developed on the basis of fluorescence properties of HS still need extensive validation on a much larger and differentiated number of HS samples before they can be considered fully applicable.

Table 1 Wavelengths of fluorescence emission and excitation maxima of potential molecular component contributors to the fluorescence of humic substances

Fluorescing molecule	*Structural formula*	*λ_{em} max (nm)*	*λ_{exc} max (nm)*
Methyl salicylate		448	302,366
Salicylic acid		410	314
3-Hydroxybenzoic acid		423	314
Protocatechuic acid (ionized)		455	340–370
3-Hydroxycinnamic acid		407	310
Caffeic acid		450	365
Ferulic acid		440	350
β-Naphtol (ionized)		460	350
Coumarins			
Unsubstituted		454	376
Hydroxy- and methoxy-		400–475	320–343
Esculetin		475	390
Scopoletin		460	390
Others disubstituted		430–462	350–419
Chromone derivatives		409–490	320–346
Xanthone and hydroxyxanthones		456	410
		465	343, 365
Flavones and isoflavones		415–475	313–365
Hydroxyquinolines		450	350
Schiff-base derivatives	 $-N=\overset{\vert}{C}-\overset{\vert}{C}=\overset{\vert}{C}-N-$	470	360–390

Adapted from Senesi N, Miano TM, Provenzano MR, and Brunetti G (1991) Characterization, differentiation and classification of humic substances by fluorescence spectroscopy. *Soil Science* 152: 259–271.

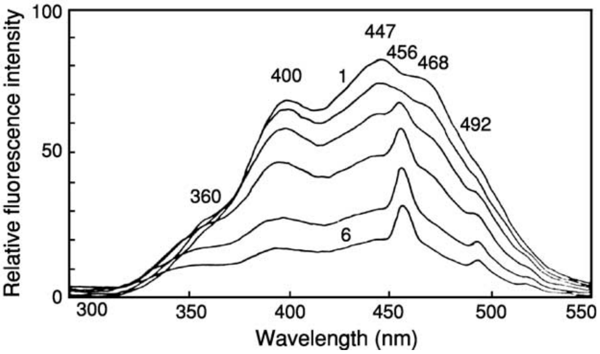


Figure 5 Effect of concentration (line 1, 120; 2, 100; 3, 75; 4, 50; 5, 25; 6, 10 mg l^{-1}) on the fluorescence synchronous-scan spectrum ($\Delta\lambda = \lambda_{em} - \lambda_{exc} = 18$ nm) of a soil fulvic acid. Reproduced with permission from Senesi N (1990) Molecular and quantitative aspects of the chemistry of fulvic acid and its interactions with metal ions and organic chemicals. Part II. The fluorescence spectroscopy approach. *Analisis Chimica Acta* 232: 77–106.

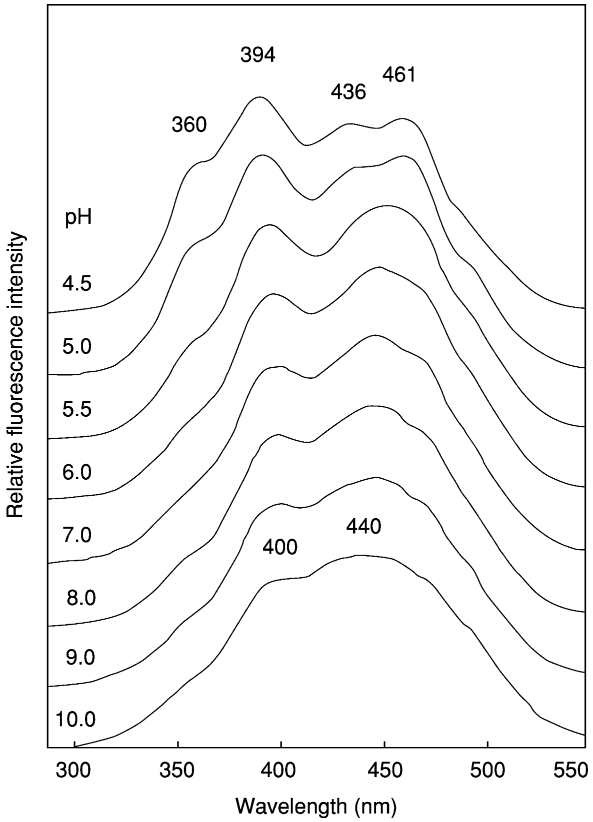


Figure 6 Effect of pH on the fluorescence synchronous-scan spectrum ($\Delta\lambda = \lambda_{em} - \lambda_{exc} = 18$ nm) of a soil fulvic acid. Reproduced with permission from Senesi N (1990) Molecular and quantitative aspects of the chemistry of fulvic acid and its interactions with metal ions and organic chemicals. Part II. The fluorescence spectroscopy approach. *Analisis Chimica Acta* 232: 77–106.

Fluorescence Efficiency and Lifetime

A fluorescence decay curve not fitting a single exponential function has been obtained for soil FAs; this indicates a range of fluorescence lifetimes as it would be expected for a mixture of fluorophores. Fluorescence efficiencies of about 0.1 are estimated for FAs from experimental lifetime data of the order of 15–20 ns. However, much smaller fluorescence efficiencies, ranging from 0.0017 to 0.0006, are measured by comparison of emission intensities of FA with quinine sulfate, which suggests the presence of only about 1% of efficient fluorophores in FAs.

The small fluorescence efficiency of FAs can also be attributed to quenching effects caused by the presence of paramagnetic metal ions and/or to a significant presence of free radicals and to other factors such as molecular size and degree of oxidation. Furthermore, fluorescence efficiency of FAs decreases with increasing MW.

Conformational Studies by Fluorescence Polarization

Fluorescence polarization is an attractive method for studying the conformation of HS because it does not require any chemical modification of the molecule, and allows measurements at concentrations close to those naturally occurring in aqueous media. The rotational relaxation time (RRT) of soil FAs obtained by this method does not change over the pH range 5–8, or at FA concentrations from 3.3×10^{-5} to 3.3×10^{-4} mol l^{-1}, or on varying the ionic strength of the medium, thus suggesting no aggregation or configuration change for FA in such conditions. The net RRT value of 2.0 ns calculated for FA based on a fluorescence lifetime of 2.1 ns is in favor of a flat extended conformation for FA. If the conformation of FA were cylindrical or rod-like, a much shorter RRT would be expected.

Values of RRT measured by time-dependent fluorescence depolarization for a soil FA in glycerol–water (40/60, w/w) solutions indicate that, at small ionic strength and FA concentration, disaggregation and molecular expansion occur for FA with increasing pH from 2.0 to 8.0, which is possibly due to the accumulation of negative charge from increasing carboxylate ionization. In contrast, RRT data obtained at low pH (2.0–4.0) and/or large FA concentration at large ionic strength are indicative of the existence of FA in an elongated conformation and of its tendency to aggregate, probably through the hydrogen bonding formation favored by the large presence of protonated acidic groups. These results confirm that FA behaves like a flexible polyelectrolyte in solutions approximating naturally occurring

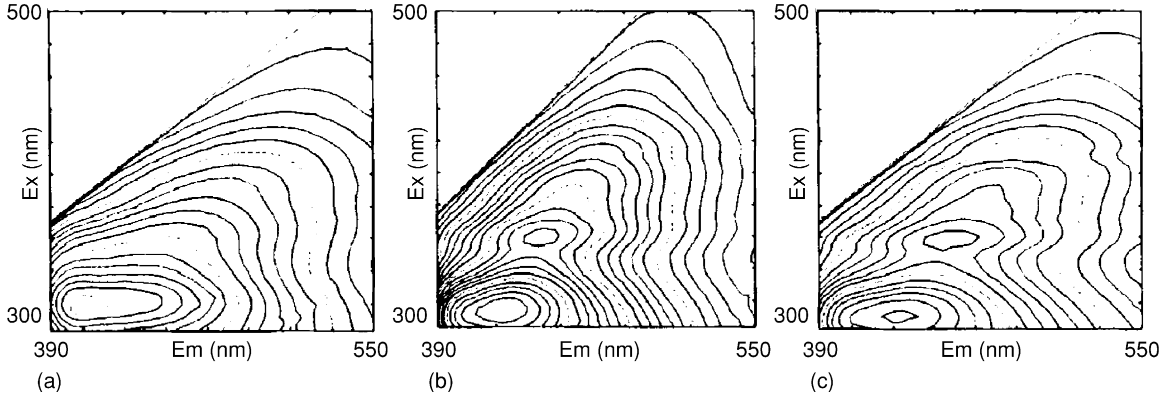


Figure 7 Absorbance-corrected excitation–emission matrix spectra of a soil fulvic acid (50 mg l^{-1}, 0.010 mol l^{-1} KCl) at pH 2.0 (a), 6.0 (b), and 10.0 (c). Reproduced with permission from Mobed JJ, Hemmingsen SL, Autry JL, and McGown LB (1996) Fluorescence characterization of IHSS humic substances: total luminescence spectra with absorbance correction. *Environmental Science and Technology* 30: 3061–3065.

concentrations. This information is very important in the solution chemistry of FA, as its structure and reactivity are markedly influenced by the conformation, i.e., molecular size and shape.

Diffusion Coefficients Determined by Fluorescence Correlation Spectroscopy

Average diffusion coefficients in the range of $2–3 \times 10^{-10}\,\mathrm{m^2\,s^{-1}}$ are determined for FAs and HAs by the FCS method in a range of concentration, hydration time, pH, and ionic strength, which corresponds to hydrodynamic diameters of approximately 1.5–2.1 nm. The slight decrease in diffusion coefficients with decreasing pH from 10 to 3 suggests that some aggregation occurs with formation of dimers and trimers. The small or insignificant reduction of the diffusion coefficient measured with increasing ionic strength up to 100 mmol l^{-1} suggests the relative rigidity of HS molecules. No effect is observed as a result of changes in the concentration (1–50 mg l^{-1}) and hydration time (1–14 days) of HA and FA and of using a Ca^{2+} salt or an Na^+ salt at an ionic strength of 5 mmol l^{-1}. For most conditions studied, FAs show a diffusion coefficient larger than that of HAs, which is in agreement with their respective molar masses.

Fluorescence Quenching of Soil Organic Components by Metal Ions

Molecular and Mechanistic Aspects

Metal ions, especially paramagnetic ions, are able to quench the fluorescence of HS by enhancing the rate of some nonradiative processes that compete with fluorescence, such as intersystem crossing. Large evidence is obtained that the 'close' association of metal ion and organic ligand, i.e., the formation of strong metal–HS complexes, is necessary for efficient quenching. Paramagnetic metal ions, such as Cu^{2+}, Fe^{3+}, Fe^{2+}, Ni^{2+}, Cr^{3+}, and VO^{2+}, can effectively quench HS fluorescence via intramolecular energy transfer by forming strong complexes. The paramagnetic ions Co^{2+} and Mn^{2+}, and diamagnetic ions such as Pb^{2+} and Al^{3+}, which form weaker complexes with HS, show a much less pronounced quenching effect, whereas cations like Cd^{2+}, K^+, Na^+, Ca^{2+}, and Ba^{2+} show no quenching effect.

Wavelength shifts of fluorescence emission maxima and/or excitation peaks are often observed upon interaction of HS with paramagnetic metal ions. For example, complexation of Cu^{2+} and Fe^{3+} to a soil FA at pH 4 and 6 causes a shift to longer wavelength (390 nm) of the 360-nm excitation peak of unreacted FA. In contrast, the excitation peaks of FA are shifted toward shorter wavelengths by FA interaction with Sn^{2+} more than with Ni^{2+}.

The fluorescence quenching of various soil FAs occurring upon complexation with Mn^{2+} ions decreases as the MW of FA increases. Synchronous-scan fluorescence analysis of three different MW fractions of a soil FA shows that the affinity for Al^{3+} and UO_2^{2+} ions of the large-MW fraction is larger than that of the small-MW fraction.

Fluorescence quenching of HAs and FAs of various origin by Cu^{2+}, Co^{2+}, Pb^{2+}, Ni^{2+}, and Mn^{2+} increases as pH increases above *c.* 3 for Cu^{2+} and above *c.* 4 for the other three ions (Figure 8). At pH 6 and 7, Cu^{2+} quenches nearly 80% of FA fluorescence, Co^{2+} and Ni^{2+} are less than half effective, and Mn^{2+} quenches about one-third of FA fluorescence. These findings are in favor of a static quenching mechanism depending primarily on the fraction of FA ligand sites complexing the metal ions, and exclude a dynamic quenching mechanism occurring by collision of FA fluorophores with the metal ions. The greater quenching ability of Cu^{2+} is attributed to its capacity

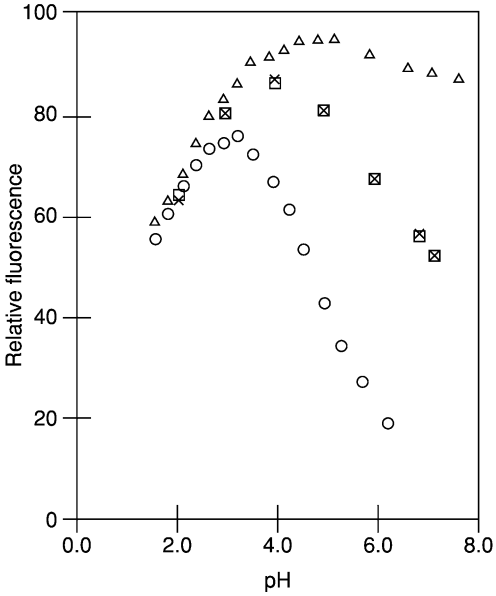


Figure 8 Fluorescence of soil fulvic acid solutions (5×10^{-5} mol l^{-1}) as a function of pH with no metal ion added (triangles) and with threefold molar excess of Cu^{2+} (circles), Co^{2+} (crosses), or Ni^{2+} (squares). Adapted from Saar RA and Weber JH (1980) Comparison of spectrofluorimetry and ion-selective electrode potentiometry for determination of complexes between fulvic acid and heavy-metal ions. *Analytical Chemistry* 52: 2095–2100.

to form strong inner-sphere complexes with both strongly acidic carboxyl and weakly acidic phenolic groups in salicylic and phtalic acid-type sites. In contrast, Mn^{2+}, and likely Co^{2+}, would form outer-sphere complexes with FA not involving weakly acidic sites.

Fluorescence quenching kinetics studies of a soil FA added with increasing amounts of Cu^{2+} suggest the occurrence of two mechanisms for binding of Cu^{2+} to FA. During the initial phase, when the Cu^{2+}-to-FA ratio increases up to 0.6, 60% of the total fluorescence quenching and a pH change from 6.4 to 5.8 are measured, indicating the occurrence of a very rapid protolytic reaction (strictly complexation). During the second phase, from a Cu^{2+}-to-FA ratio of 0.6–1.0, the reaction is slower and pH remains relatively stable, thus suggesting charge neutralization as the dominant process. Above a Cu^{2+}-to-FA ratio of 1.0, fluorescence remains almost constant, indicating that Cu^{2+} is no longer specifically bound.

Synchronous-scan fluorescence is a more revealing measurement than emission fluorescence for studying quenching of FA fluorescence by metal ions. Principal component analysis of synchronous quenching spectra obtained by difference before-and-after additions of Cu^{2+}, Pb^{2+}, Ni^{2+}, Co^{2+}, Ni^{2+}, Mn^{2+}, Mg^{2+}, and Al^{3+} at various total concentrations indicates that multiple fluorescent binding sites of FA are involved in divalent metal complexation in solutions at pH 7.5, but not at pH 5. The ion Al^{3+} is shown to compete with Cu^{2+} for FA binding sites, and to displace Mg^{2+} bound to FA. Correlation of synchronous fluorescence and time-resolved fluorescence measurements suggests that a three-component model for FA accommodates both pH dependence and metal ion quenching.

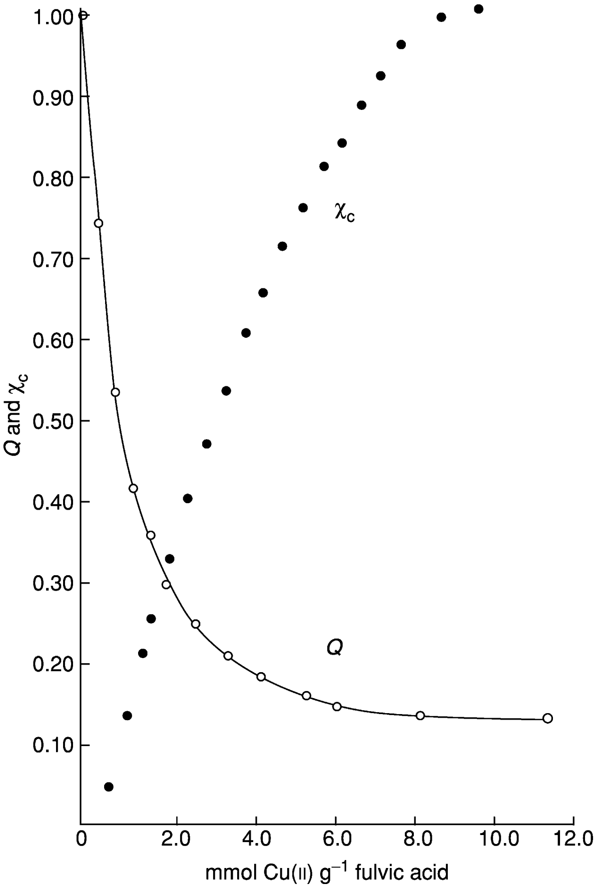


Figure 9 Comparsion of two experimental probes of metal binding to a soil fulvic acid. χ_c, increasing fractions of fulvic acid binding sites occupied by Cu^{2+} as calculated from ion-specific electrode titration; *Q*, relative fluorescence quenching of fulvic acid. Adpated from Underdown AW, Langford CH, and Gamble DS (1981) The fluorescence and visible absorbance of Cu(II) and Mn(II) complexes of fulvic acid: the effect of metal ion loading. *Canadian Journal of Soil Science* 61: 469–474.

A close inverse relationship exists between the relative fluorescence quenching, *Q*, measured at the maximum emission intensity for a soil FA titrated with Cu^{2+} ions and the fraction of binding sites, χ_c, occupied by the metal ion obtained by Cu^{2+} ion-specific electrode (ISE) potentiometry (Figure 9). The quenching curve obtained for Mn^{2+} has similar characteristics but Mn^{2+}, which is about 100 times more weakly complexed than Cu^{2+}, requires about 100 times the concentration of Cu^{2+} to yield the same degree of quenching. These results confirm the occurrence of a static, and not collisional quenching mechanism.

Comparative studies by ISE potentiometry, anodic stripping voltammetry (ASV), and fluorescence quenching show that no single site describes adequately Cu^{2+}–FA complexation over the entire pH range, but a mixed mode of coordination would occur, with the dominant binding sites varying with pH and metal-to-ligand ratio. Aromatic amino acidic groups, such as tyrosine and phenylamine moieties, and citrate and malonate groups in FA, are suggested to contribute to Cu^{2+} binding from pH 3 to 7, whereas at higher pHs, polydentate groups such as salicylates and phtalates would be mainly involved in Cu^{2+} complexation by FA.

Quantitative Aspects

Fluorescence quenching titrations of a soil FA with Cu^{2+} have provided a complexing capacity (C_L) value of FA for about 20 mol l^{-1} Cu^{2+} at pH 6 and 7, and a conditional stability constant (K_b) of the complex FA–Cu^{2+} two to three times greater at pH 7 than at pH 5 and 6. Average C_L and K_b values determined by fluorescence quenching are of the same order of magnitude as those determined by polarographic titration. At pH 7 and 6 the final fluorescence intensity is approximately 20% of the initial value for FA, and it is 40% at pH 5, which indicates that most FA fluorophores are directly involved in Cu^{2+} complexation.

With increasing Cu^{2+} concentration, the fluorescence intensity of the shorter EEM wavelength pair of a soil FA decreases more than that of the longer-wavelength pair. The K_b value evaluated assuming a 1:1 complexation stoichiometry for Cu^{2+}-FA binding is slightly larger for the shorter EEM wavelength pair than for the longer one, thus confirming the existence of different types of metal-binding sites for Cu^{2+} in FA.

Synchronous-scan fluorescence spectra of FAs added with increasing amounts of Be^{2+} at pH values from 4 to 7 indicate the presence of a fluorescent component not affected by Be^{2+} and another component resulting from the formation of relatively strong fluorescent complexes with Be^{2+} by different FA binding sites. The calculated C_L are smaller at pH 4 than at higher pH, as expected by protonation of some FA binding sites that become unavailable to complexation at low pH. The K_b values also increase from pH 4 up to pH 6, but decrease slightly at pH 7. A similar analysis applied to complexation of Cu^{2+}, UO_2^{2+}, and Al^{3+} by FA at various pH values and concentrations shows, in the case of Al^{3+}, the presence of one component corresponding to the formation of a fluorescent complex and another component corresponding to a nonfluorescent complex between Al^{3+} and FA. Calculated K_b values indicate that the stability of FA–Al^{3+} complexes decreases markedly at pH < 4.5.

With increasing Al^{3+} concentration, the fluorescence polarization increases for the smallest MW fraction of a soil FA and decreases for the largest MW fractions, suggesting that the former FA fraction does not change and the latter fractions change conformation upon complexation with Al^{3+} ions. The K_b values for Al^{3+}–FA complexes can be calculated using the anisotropy values obtained from polarization data. Time-resolved laser fluorescence spectroscopy has also been used to calculate K_b values of complexes formed by soil HAs with trivalent metal ions such as Cm^{3+} and Eu^{3+}.

In general, the information obtained by fluorescence quenching on metal–HS complexation is comparable to that obtained by other techniques, including ASV, ISE potentiometry, and dialysis/atomic absorption spectrometry. Although several limitations and assumptions are inherent to fluorescence data analysis, this method has several advantages over other methods in that it is relatively rapid, no separation is required between bound and free metal ion, there is no need to add supporting electrolyte or buffer or adsorbing material to the sample, neither CO_2 nor O_2 interferes with the measurement, and its sensitivity is enough for application to unmodified, natural organic ligands without preconcentration. A combined ISE potentiometry–fluorescence spectroscopy approach shows the additional advantages of avoidance of errors due to inorganic metal complexes, and applicability to environmental matrices where the metal ISE may not be suitable and to paramagnetic ions for which no ISE is available. The major disadvantage of fluorescence spectroscopy, that is, its efficiency only with strongly binding paramagnetic metal ions such as Cu^{2+}, can be, however, overcome by using fluorescent probes such as the lantanide ion probe, which is particularly promising for studies at environmental concentrations of both the metal and HS.

Sorption/Binding of Fluorescent Organic Pollutants to Soil Organic Components

Both fluorescence polarization and fluorescence quenching methods can be applied with success for molecular and quantitative studies of the binding/sorption of organic pollutants to soil organic components, and especially HS. The major advantage of these methods is that no separation step is required, which eliminates possible errors due to incomplete separation of the free/dissolved from the bound/

sorbed pollutant with consequent potential improvements in reproducibility. Furthermore, both methods are relatively rapid and convenient, and offer good precision and great inherent sensitivity, thus enabling measurements of pollutant concentration at environmental levels.

The fluorescence polarization method is preferentially applied where the pollutant molecule is an efficient fluorophore with respect to the sorbent, i.e., the HS molecule. Alternatively, the fluorescence quenching method can be applied to systems where the fluorescence of the pollutant is affected by that of HS.

Fluorescence Polarization

The fluorescence of perylene in glycerol–water (75/25, w/w) solutions at pH values of 2, 5, and 11 becomes more polarized as a function of added FA (Figure 10), which indicates that perylene binds to FA, forming a larger species that rotates more slowly in solution than does perylene alone. Polarization reaches a maximum value and then remains constant at FA levels above *c.* $3 \times 10^{-6}\,\mathrm{mol\,l^{-1}}$, which suggests that binding is complete at this concentration. The relatively small overall change in polarization suggests that the reaction involves one perylene molecule per one FA molecule, i.e., a 1:1 stoichiometry.

Based on these assumptions, a binding equilibrium constant, K_b, can be calculated for the perylene-FA reaction:

$$\text{pery} + \text{FA} = \text{pery} - \text{FA} \qquad [12]$$

directly from polarization data by derivation of eqn [11], where the Φ_f/Φ_b term can be considered equal to 1:

$$K_b = [\text{pery} - \text{FA}]/[\text{pery}][\text{FA}] = (p - p_f)/\{(p_b - p)\{C_{FA} - [(p - p_f)/(p_b - p)]C_{pery}\}\} \qquad [13]$$

where p_f and p_b are the initial (no FA added) and final polarization values of perylene, and C_{FA} and C_{pery} are the total concentrations of FA and perylene, respectively. The K_b calculated from eqn [13] are 1.2×10^6, 1.8×10^6, and 1.5×10^6 at pH 2, 7, and 11, respectively. These values, which are expected to be considerably larger in water, confirm the existence of extended hydrophobic surface regions in FA available for association to hydrophobic organic pollutants such as perylene and other polycyclic aromatic hydrocarbons (PAHs), and suggest that this process is likely to be significant in natural systems.

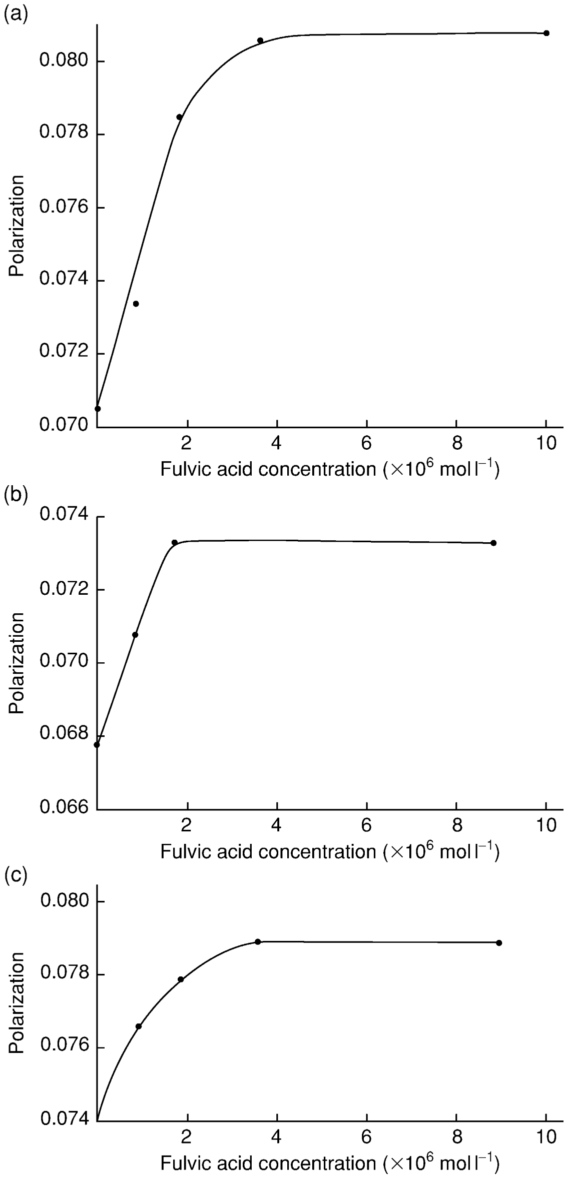


Figure 10 Perylene polarization as a function of added fulvic acid in glycerol–water (75/25, w/w) solutions of perylene ($4.9 \times 10^{-7}\,\mathrm{mol\,l^{-1}}$) at pH 2.0 (a), 7.0 (b), and 11.0 (c). Reproduced with permission from Roemelt PM and Seitz WR (1982) Fluorescence polarization studies of perylene–fulvic acid binding. *Environmental Science and Technology* 16: 613–616.

The fluorescence polarization method can be extended to nonfluorescent compounds by either measuring the changes of intrinsic fluorescence polarization of the sorbent HS molecule or performing competitive binding experiments using an adequate fluorophore probe. In the first case, the polarization changes are likely to be small and difficult to measure with precision, thus good-quality instrumentation is required, whereas in the second case, irreversible binding and slow kinetics can complicate the experiment.

Fluorescence Quenching

The rate of binding of benzo(*a*)pyrene (BaP) to a water-dissolved HA at three concentrations has been analyzed by measuring the quenching of BaP fluorescence, i.e., the increase in F_0/F ratio (eqn [9]). At any concentration, quenching is complete within 5–10 min of HA addition (**Figure 11**), which suggests a rapid association of BaP to HA and the existence of only a fast kinetic component.

The K_b values for the association of various PAHs, including BaP, perylene, fenantrene, and anthracene, with water-dissolved HAs and FAs can be derived using the linear Stern–Volmer plots (eqn [9]) obtained by measuring the fractional decrease of fluorescence intensity, F_0/F, as a function of the added HS quencher. Further analysis of fluorescence data indicates the occurrence of a static quenching process, i.e., the formation of a nonfluorescent complex between PAH and HS. The fluorescence quenching method also allows the measurement of the partition coefficients and adsorption coefficients normalized with respect to organic carbon content, K_{OC}, of compounds with large fluorescence efficiencies, such as several PAHs, to soil HAs and FAs under various pH and ionic strength conditions.

Fluorescence quenching and synchronous-scan fluorescence spectroscopy have been used to obtain conditional stability constants for the binding of carbamate pesticides to soil DOM and HA. For example, at pH 6, the binding constant of carbofuran is greater than that of carbaryl and aldicarb. These results can be used to predict the potential transport of pesticides down the soil profile.

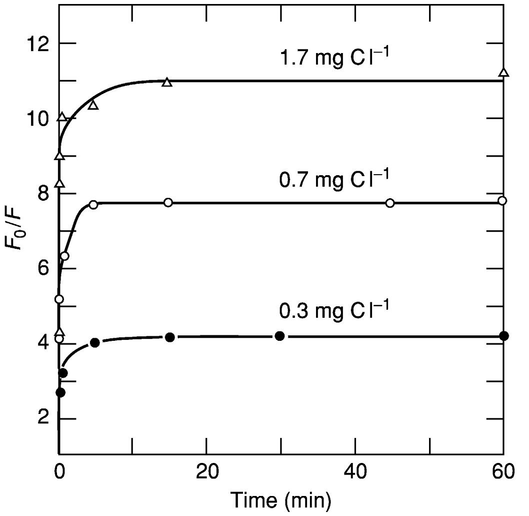


Figure 11 Time course of quenching benzo(*a*)pyrene fluorescence following the addition of humic acid (HA) at various concentrations at time zero. F_0/F is the ratio of fluorescence in the absence of HA to that in the presence of HA. Reproduced with permission from McCarthy JF and Jimenez BD (1985) Interaction between polycyclic aromatic hydrocarbons and dissolved humic material: binding and dissociation. *Environmental Science and Technology* 19: 1072–1076.

The association index (AI), formulated as:

$$\mathrm{AI} = (r_e - r_m)/r_e \qquad [14]$$

where r_e and r_m are, respectively, the hypothetical or 'expected' fluorescence anisotropy value in the absence of interactions and the experimental value of anisotropy, has been used as a quantitative parameter to measure the degree of association between various HAs and the 2,5-diphenyloxazole fluorescent probe as a function of the HA nature, HA and probe concentrations, pH, and ionic strength. Evidence has been obtained of the existence of HA pseudomicelles that are able to sequester hydrophobic species at an extent that is a function of the nature and distribution of HA functional groups, the size and flexibility of HA, and solution polydispersity and conditions. This index has been proposed as a convenient indicator of the general tendency of HAs to associate with hydrophobic species.

Soil Mineral Components

Static and dynamic fluorescence spectroscopy of organic probes sorbed on soil clay minerals can provide important and unique insights into several aspects of specific interfacial and surface behavior of clay particles in dilute colloidal dispersions/suspensions. A probe molecule is selected so that it has a great fluorescence efficiency, is sensitive at small surface loadings, and thus suitable for dilute sample suspensions, and possesses fluorescence properties depending on environmental factors such as pH and degree of molecular aggregation. Furthermore, the adsorption mechanism of the probe by the surface of interest and the specific aspects of the interface/surface that are to be probed must be considered. Common fluorescent probes include some organic cationic dyes, such as acridine orange, methylene blue, proflavine, and R6G, and various PHAs, such as anthracene, 1,2,5,6-dibenzoanthracene, perylene, BaP, pyrene, and various cationic pyrenyl derivatives.

Advantages of the fluorescence method are:

1. The possibility to study aqueous colloidal suspensions noninvasively, i.e., without the need to stop the reaction or separate the solid and liquid phases.
2. The great analytical sensitivity that makes it possible to detect very small concentrations of

sorbed probes in hydrated clay suspensions in conditions approaching those existing in natural environments.

3. The ability of one probe species to provide multiple types of information on the colloid–solution interface under study.
4. The possibility of exploring reaction rates continuously across wide time scales ranging from minutes to microseconds.
5. The study of events occurring on nanosecond time scales using time-resolved fluorescence decay measurements.

However, limitations in using fluorescence spectroscopy to study mineral systems include excessive light scattering by dense colloidal suspensions and the attenuation of fluorescence of probe species sorbed at the mineral–solution interface by quenching cations, mostly Fe^{3+} ion located within the crystal structure or on the surface of clay minerals. The use of fluorescence is therefore restricted to suspensions of clays 'Fe-free' or containing relatively small amounts of structural Fe(III).

Structural Quenching

The fluorescent probes *tris*(2,2′-bipyridine) Ru(III) and *tris*(2,2′-bipyridine) Cr(III) are more strongly quenched by absorption on to Na-saturated Wyoming montmorillonite than on to a synthetic montmorillonite containing Cr(III) in the octahedral layer and hectorite. Furthermore, Fe(III) and Cr(III) in octahedral positions are more effective quenchers than Fe^{3+} and Cr^{3+} residing on the interlayer exchange sites. Similarly, octahedral Cu(II) contained in a synthetic hectorite exhibits a greater quenching efficiency than exchangeable Cu^{2+} for the adsorbed fluorophore [4-(1-pyrenyl)butyl] trimethylammonium (PN4). These effects can be ascribed to the close association of the adsorbed fluorophore with the siloxane ditrigonal cavities on the clay surfaces.

Fluorescence quenching can also be used to determine the location of a fluorescent species at the mineral–water interface and its partitioning into mineral micelles. For example, the inorganic fluorophore Ce^{3+} is shown to occupy preferentially ion exchange sites within the tubules of imogolite. The fluorescent probe pyrene is readily quenched by both polar and nonpolar quenchers when it is located at the water–clay micelle interface, whereas it is not quenched by a polar quencher such as Cu^{2+} when it resides in the nonpolar interior of the micelle. However, pyrene fluorescence can be readily quenched by a nonpolar quencher that can also be dissolved within the interior of the micelle.

Adsorption of Organic Chemicals to Mineral Surfaces

Fluorescence polarization has been used to study the adsorption of several PAHs, including anthracene, pyrene, perylene, 1,2,5,6-dibenzoanthracene, BaP, and chrysene, on to aqueous colloidal kaolin. In all cases a distinct decrease in fluorescence anisotropy is observed as a function of PAH concentration, which implies an RRT considerably greater at small than at large concentrations. These results suggest that, at small concentrations, most PAH molecules are adsorbed on colloidal kaolin, and their motion in solution is hindered. With increasing concentration the surface saturation of the colloid is reached, and additional molecules are no longer adsorbed and remain 'free' to rotate, leading to the decrease in fluorescence anisotropy. Furthermore, adsorption kinetics studies indicate that adsorption occurs within the first few minutes of contact.

Electrofluorescence polarization spectroscopy has been used to measure the polarization of aqueous suspensions of sepiolite, attapulgite, and hectorite particles tagged with a range of dye molecules of differing sizes and charges and oriented in an applied electric field. Data obtained make it possible to infer the different possible binding sites of the dye to the clay. For example, cationic dyes are bound predominantly within the channels of sepiolite rods, whereas no effect is observed with either attapulgite or hectorite.

Surface Acidity

Fluorophores that exhibit pH-dependent fluorescence can be used to probe clay surface acidities and protonation of organic molecules at the clay–solution interface. For example, the emission spectra of quinoline on to Na-hectorite indicate the presence of surface-adsorbed quinolinium ions in the pH range 3–9. The protonation of quinoline can be ascribed to specific acidic sites at the mineral–water interface and/or to the strong ion exchange capacity of quinolinium ions. Furthermore, characteristic emission spectra of diprotonated proflavine are observed when the monoprotonated form is added to suspensions of Na-saturated montmorillonite, Wyoming bentonite, or barasym.

In some cases, fluorescence spectra can be used to infer the presence of both protonated and unprotonated species on the clay surface. For example, aminopyrene sorbed to colloidal silica exhibits both the largely structured emission spectrum of protonated aminopyrene and the broad, featureless emission of neutral aminopyrene, which indicate the presence on the surface of both species, the protonated one on the

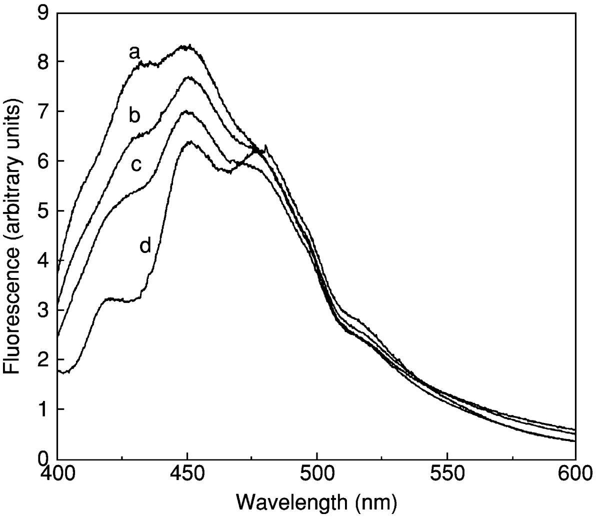


Figure 12 Effect of concentration on the emission spectra of acridine (a, 16 μmol l^{-1}; b, 65 μmol l^{-1}; c, 99 μmol l^{-1}; d, 122 μmol l^{-1}) on hectorite suspension (115 mg l^{-1}) at pH 8.8. Reproduced with permission from Traina SJ and Chattopadhyay S (1996) Spectroscopic methods of characterization surface-sorbed organic contaminants: applications of optical, electronic measurements. In: Sahwney B (ed.) *Organic Pollutant in the Environment*, vol. 8, pp. 173–197. Boulder, CO: Clay Minerals Society.

more acidic sites, and the neutral one on the less acidic sites. Fluorescence emission spectra of acridine sorbed on to Na-hectorite at a pH value well in excess of acridine pK_b (e.g., at pH 8.8) (Figure 12) show that, at very small acridine loadings, the dominant surface species is the acridinium ion, which indicates protonation promoted by the surface even at pH 8.8. At large acridine loadings, the two main components, $\tau_1 = 40$ ns and $\tau_1 = 9$ ns, of the fluorescence decay spectra of acridine adsorbed on hectorite would indicate sorption of both the acridinium ion and neutral acridine, possibly by aggregation of neutral acridine molecules on the surfaces of adsorbed acridinium ions. Regardless of the total acridine concentration, suspensions at low pH values (e.g., 4.5) result in spectra of the acridinium cation only.

Polarity of Mineral Surfaces

The great sensitivity of some fluorescent probes to the surrounding solvent polarity can provide information on the presence of polar and nonpolar regions on mineral colloid surfaces. The significantly smaller value of the ratio of fluorescence emission intensities at $\lambda = 372$ to $\lambda = 391$ nm (I/III ratio) of pyrene adsorbed to an air-dried Na-saturated laponite, with respect to that of aqueous pyrene solutions, indicates that, at small moisture contents, pyrene can be sorbed to low-polarity regions of the clay siloxanic basal planes. However, further dehydration of the clay causes an increase of the I/III ratio above the value of aqueous pyrene, which indicates the existence of an interfacial region that is more polar than bulk water.

The large I/III emission ratios of pyrene, [3-(1-pyrenyl)propyl]trimethyl ammonium (PN3), and 8-(1-pyrenyl)octyl ammonium (PN8), adsorbed on to laponite from water and methanol solutions, confirm that the charged regions of the interlamellar surfaces of clay particles are largely polar. In contrast, the very small I/III ratios observed when the fluorescent probes are sorbed on to clays coated with alkylammonium surfactants indicate that they reside in less polar environments in forms strongly associated to the adsorbed surfactants.

Organic dye molecules can also alter the apparent polarity of clay mineral surfaces. For example, the comparative analysis of the I/III emission ratios of aqueous pyrene alone and added with quinoline in the absence or in the presence of clay particles indicates the direct sorption of pyrene molecular clusters to quinolinium-coated hectorite surface.

Molecular Distribution, Conformation, Aggregation, and Mobility

Excitation spectra and fluorescence lifetime measurements show that adsorbed PN4, PN3, and PN8 cationic probes tend to cluster on the surfaces of hydrated clay minerals with formation of fluorescent dimers, i.e., excimers. The time-dependent decrease of excimer emission from the PN3 probe, which is observed upon its addition to aqueous suspensions of Na-saturated hectorite, laponite, or montmorillonite, is attributed to the initial adsorption of the probe to exterior, nonhomogeneous clay sites in the form of aggregates or hydrophobic micelles exhibiting large excimer fluorescence, followed by migration of the probe to interlayer sites with a decrease in the total amount of fluorescent excimers. The addition of Ca^{2+} ions to the above-described systems increases the excimer fluorescence intensity immediately after adsorption by inducing flocculation of clay particles and increasing the concentration of PN3 probes on the exterior surfaces. Subsequently, PN3 cations migrate into the interlayers of the more flocculated Ca^{2+}-clays, thus decreasing their mobility and excimer emission intensity.

The distribution of pyrenyl cations on clay surfaces can be altered by the presence of other organic cations such as alkylammonium surfactants. For example, PN4 cations can be 'solubilized' into the surfactant hemimicelles and form clusters that are dispersed and more homogeneously distributed on the clay surface. This results in increased PN4 mobility and the corresponding reduction of excimer emission.

Increases of the chain length of the alkylammonium surfactant produce a further decrease in excimer emission.

The effect of pH on the interfacial conformations of polymeric, polyelectrolytic macromolecules, such as pyrene-labeled polyacrylic acid (Py-PAA), can be studied by analyzing the ratio of the Py-PAA monomer emission to the excimer emission both in solution and adsorbed to alumina in aqueous suspension. When Py-PAA is adsorbed on to alumina at pH 8, and the solid is separated from the aqueous phase and then resuspended in a pH 4.8 solution, no change in the excimer emission intensity is observed. This result suggests that the polymer is adsorbed in the uncoiled, linear conformation and remains uncoiled independent of the pH value of the surrounding medium, i.e., no conformational change occurs as a function of pH. In contrast, when Py-PAA is adsorbed on to alumina at pH 4.7, and the particles are separated and then resuspended at pH 8, a decrease in excimer emission intensity is measured. This indicates the occurrence of some uncoiling of the adsorbed polymers. This means that coiled polymers loosely bound to alumina surface at low pH can undergo some conformational reorganization when exposed to a higher pH.

Fluorescence emission spectroscopy shows great promise for studying the molecular organization of other polymers, such as HAs and FAs, at interfacial mineral–solution regions. However, confirmation of fluorescence results is generally required by the use of one or more additional techniques, such as NMR.

Biochemical Transformations on Clay Surfaces

Bacterial degradation of quinoline in suspensions of hectorite has been studied by fluorescence spectroscopy. At pH values well above the pK_b of quinoline, the only fluorescent species in the system is the protonated quinolinium ion sorbed on to hectorite surfaces. The concentration decrease of adsorbed quinolinium can be monitored during its microbial degradation together with the production of the metabolite, 2-hydroxyquinoline. Comparison of degradation rates in the presence and absence of hectorite suggests that desorption of adsorbed quinolinium ion is the rate-limiting step in the bacterial degradation of quinoline.

Ligand-Promoted Mineral Dissolution

Direct evidence can be provided for the formation of fluorescent inner-sphere complexes between Al^{3+} ions and 8-hydroxyquinoline-5-sulfonate at the Al oxide–water interface. Fluorescence measurements show that the concentration of the surface complex is constant during long-term dissolution of the oxide, which indicates that the oxide surface is continuously regenerated during dissolution; this confirms a surface-controlled process of mineral dissolution.

List of Technical Nomenclature

λ	wavelength
λ_{em}	emission wavelength
λ_{exc}	excitation wavelength
$\Delta\lambda$	wavelength difference
Φ_f, Φ_s	fluorescence efficiency, or quantum yield
τ	fluorescence lifetime
q	photon output of the source
A	absorbance
ϵ	molar absorptivity
b	path length of the cell
I	fluorescence intensity
$I_{\parallel}$	fluorescence intensity in the parallel plane
$I_{\perp}$	fluroscence intensity in the perpendicular plane
p	polarization
C	molar concentration
K_b	conditional stability constant
C_L	complexing capacity
K_{oc}	adsorption coefficient normalized to organic carbon content
Q	quenching of fluorescence quantum yield
χ_c	fraction of binding sites occupied
nm	nanometer
AI	association index
ASV	anodic stripping voltammetric
BaP	benzo(*a*)pyrene
DOM	dissolved organic matter
EEM	emission–excitation matrix
EPR	electron paramagnetic resonance
FA	fulvic acid
FCS	fluorescence correlation spectroscopy
FZI	fluorescence zone integration
HA	humic acid

HS	humic substances
IR	infrared
ISE	ion-specific electrode
MW	molecular weight
NMR	nuclear magnetic resonance
PAH	polycyclic aromatic hydrocarbon
PN3	[3-(1-pyrenyl)propyl]trimethylammonium
PN4	[4-(1-pyrenyl)butyl]trimethylammonium
PN8	8-(1-pyrenyl)octylammonium
Py-PAA	pyrene-labeled polyacrylic acid
R6G	rhodamine 6G
RRT	rotational relaxation time
UMC	uncorrected matrix correlation

See also: **Clay Minerals**; **Fourier Transform Infrared Spectroscopy**; **Heavy Metals**; **Organic Matter:** Principles and Processes; Interactions with Metals; **Pesticides**; **Sorption:** Metals; Organic Chemicals

Further Reading

Gauthier TD, Shane EC, Guerin WF, Seitz WR, and Grant CL (1986) Fluorescence quenching method for determining equilibrium constants for polycyclic aromatic hydrocarbons binding to dissolved humic materials. *Environmental Science and Technology* 20: 1162–1166.

Grant CL and Seitz WR (1982) The potential of fluorescence polarization for measuring sorption isotherms of organics. *Soil Science* 133: 289–294.

Kraayenhof R, Visser AJWG, and Wolfbeis OS (2002) *Fluorescence Spectroscopy, Imaging and Probes – New Tools in Chemical, Physical and Life Sciences*. Springer Series on Fluorescence, Methods and Applications. Heidelberg: Springer-Verlag.

Lacowicz JR (1986) *Principles of Fluorescence Spectroscopy*. New York: Plenum Press.

Mobed JJ, Hemmingsen SL, Autry JL, and McGown LB (1996) Fluorescence characterization of IHSS humic substances: total luminescence spectra with absorbance correction. *Environmental Science and Technology* 30: 3061–3065.

Roemelt PM and Seitz WR (1982) Fluorescence polarization studies of perylene–fulvic acid binding. *Environmental Science and Technology* 16: 613–616.

Schulman SG (1985) *Molecular Luminescence Spectroscopy. Methods and Applications*. New York: Wiley.

Senesi N (1990) Molecular and quantitative aspects of the chemistry of fulvic acid and its interactions with metal ions and organic chemicals. Part II. The fluorescence spectroscopy approach. *Analisis Chimica Acta* 232: 77–106.

Senesi N and Miano TM (1993) Potential of fluorescence spectroscopy in the study of interactions of pesticides with natural organic matter. In: Mansour M (ed.) *Fate and Prediction of Environmental Chemicals in Soils, Plants and Aquatic Systems*, pp. 111–121. Boca Raton, FL: Lewis Publishes.

Senesi N, Miano TM, Provenzano MR, and Brunetti G (1991) Characterization, differentiation, and classification of humic substances by fluorescence spectroscopy. *Soil Science* 152: 259–271.

Sharma A and Schulman SG (1999) *Introduction to Fluorescence Spectroscopy*. New York: Wiley.

Traina SJ (1990) Application of luminescence spectroscopy to studies of colloid–solution interfaces. In: Stewart BA (ed.) *Advances in Soil Science*, vol. 14, pp. 167–190. New York: Springer-Verlag.

Traina SJ and Chattopadhyay S (1996) Spectroscopic methods for characterizing surface-sorbed organic contaminants: applications of optical, electronic measurements. In: Sahwney B (ed.) *Organic Pollutant in the Environment*, vol. 8, pp. 173–197. Boulder, CO: Clay Minerals Society.

Valeur B, Brochon J-C, and Wolfbeis OS (2001) *New Trends in Fluorescence Spectroscopy – Applications to Chemical and Life Science*. Springer Series on Fluorescence, Methods and Applications. Heidelberg: Springer-Verlag.

FOLIAR APPLICATIONS OF NUTRIENTS

M Tagliavini and M Toselli, University of Bologna, Bologna, Italy

Introduction

It is well known that certain nutrients in gaseous form are absorbed by plant leaves. While leaves have not developed as extensively as roots as absorbing organs, in the evolution from aquatic to terrestrial plants, they have retained the ability to take up water-soluble nutrients. The nutrition of plants through aboveground organs is generally regarded as a technique to supply nutrients quickly to a target organ. This feature is particularly important for agricultural crops, which perform best when their nutrient demand is fully satisfied throughout the growth cycle.

Since most foliar fertilizers can be mixed with pesticides in the sprayer tank, foliar fertilization is often part of routine cultural practices to complement the soil-nutrient supply. In addition, for many nutrients and under specific circumstances, foliar fertilization provides nutrients to the plant with greater efficiency than soil fertilization and hence represents an important alternative to the latter. Foliar application of nutrients is thus seen as a promising technique in reducing the amount of chemicals in the environment in promoting sustainable agriculture. Several variables influence the effectiveness of foliar-applied fertilizers and may explain the often-reported inconsistency in plant response to foliar nutrition.

Efficiency of Foliar-Applied Fertilizers

In order to produce a positive effect on the mineral nutrition of plants, several requisites must be met:

1. The applied nutrients must be in an available form for leaf absorption;
2. The applied nutrients must reach the target organ (often leaves but also fruits);
3. The target organ should have a sufficient retention area so little is dispersed in the environment;
4. The internal leaf structure, environmental conditions, and nutrient chemistry should be such that nutrients either in ionic or in nonionic form can penetrate the leaf cuticle and enter the apoplast of the cells of the target organ;
5. The given nutrient should be transportable to the target tissues if different from the intercepted organ.

Meeting these requirements, however, does not necessarily improve plant performance. Positive responses to foliar as well as to soil fertilization is expected only when there is a need for any given nutrient: if nutrient availability exceeds nutrient demand, no response to foliar fertilizers is likely.

There are a number of situations in which foliar nutrition can effectively replace or complement (Table 1) soil fertilization. Because the amount of nutrients to be supplied through the leaves is limited by the fact that, above a certain threshold concentration, foliar sprays can be toxic, foliar fertilization can meet plant nutrient demand if the requirement for a given nutrient is low, as in the case of microelements (boron, copper, iron, manganese, molybdenum, and zinc) or when soils already have adequate fertility, so that only small additional amounts are needed. Multiple foliar applications may be needed to ensure adequate nutrient delivery.

There is general agreement that foliar fertilization isan efficient method to prevent or overcome transient nutrient deficiency over the vegetative and

Table 1 Major agronomic and physiology-related conditions promoting the effectiveness of foliar nutrient supply

Objective	*Example*
Prevent or cure transient nutrient deficiency	When nutrient demand exceeds root uptake rate
Bypass soil conditions limiting nutrient availability or uptake and accelerate plant response	Unfavorable pH and conditions for root growth, including soil temperature, moisture, and aeration
Increase the buildup of nutrient reserves for storage and successive remobilization the following year	It applies to perennial plants and to nutrients that are phloem-mobile such as nitrogen
Cure or prevent nutrient deficiencies of specific organs that occur despite optimal root uptake due to problems in within-plant distribution of a given nutrient	It applies to nutrients such as Ca which move mainly through the transpiration stream toward the leaves, leading to calcium-related disorders in fruits, e.g., in apple and in tomato
Supply nutrients to plants with deep root systems when soil-surface application of fertilizers is almost without effect	It applies to trees, especially for nutrients such as K, Ca, or Mg, whose penetration to root zone layer is sometimes difficult and in general to conditions of absence of irrigation or rain

reproductive season for both macro- and micro-nutrients. These situations occur when demand temporarily exceeds root uptake capacity for a given nutrient. In highly productive plants, for example, during specific stages of the reproductive cycle, frequently leaves suffer from a temporary nutrient deficiency. Seeds and fruits withdraw from vegetative parts significant amounts of nitrogen, phosphorus, and potassium (with absolute amounts depending on the crop) at a rate higher than that supplied by root uptake, which often declines as a result of decreasing carbohydrate supply to the root system. Transient nutrient deficiency also occurs in perennial plants during spring, when remobilization of nutrients from storage sites is complete and root uptake is not sufficiently efficient (e.g., due to low soil temperature, excess moisture, or poor root growth) to meet plant requirements: transient deficiency for macronutrients such as nitrogen or micronutrients such as boron often compromises fruit set when occurring during anthesis.

Foliar application is often preferred to soil fertilization when soil constraints such as pH or calcium carbonate limit the availability of nutrients such as zinc, iron, and manganese or when deep-rooted perennial plants are grown in soils with high fixing capacity for nutrients such as potassium. The addition of iron salts to an alkaline or calcareous soil is usually ineffective, as soil mineral iron (often Fe oxides) is poorly available at high pH. For high-value agricultural commodities, the Fe deficiency problem is often overcome by soil application of expensive, synthetic Fe chelates, but foliar application of iron salts may be almost equally effective and represents an economical alternative where growers cannot afford Fe chelates to cure or prevent iron deficiency. It must be stressed, however, that if Fe deficiency persists for long periods, and because Fe mobility in the phloem is very low, repeated foliar applications may be needed during active stages of shoot development to satisfy the Fe requirement of developing leaves. Leaf application of zinc and manganese, often as sulfate salts, is frequently the only way to cure or prevent deficiencies in alkaline soils, because these micronutrients are rapidly fixed when applied to the soil and become unavailable to plants. If applied in high-pH soils, zinc and manganese chelates are, moreover, not stable, as in the case of ethylenediaminetetraacetic acid (EDTA) and diethylenetriaminepentaacetic acid (DTPA), or have little affinity to Mn and Zn, as ethylenediamine-di(*o*-hydroxyphenylacetic) acid (EDDHA). The rationale for preferring foliar application to soil fertilization in optimizing micronutrient nutrition also includes the fact that, in some types of soils, soil supply may increase their availability from deficiency to excess levels, producing undesired effects.

One typical example of the advantage of foliar over soil nutrient application is related to calcium (Ca) nutrition in fruit plants. Because this element moves within the plant almost entirely in the xylem sap via the transpiration stream, it tends to accumulate preferentially in organs whose transpiration rate is high, such as the leaves, with only limited partitioning to fruits. Expanding fruits have a relatively low demand for calcium, but, in some species of marked horticultural interest, low levels of Ca in the fruit can lead to Ca deficiency-related symptoms that compromise fruit quality and storage life, as exemplified by bitter pit in apple and black end-rot in tomato. As these disorders also occur in soils well supplied with Ca and in situations where adequate concentrations of Ca are found in leaves, soil applications are of generally little or no value, and prevention and cure have to rely on direct application of Ca-based products on the fruit.

Application of Foliar Nutrients

The amount of nutrients that can enter the plant through the foliage is a function of the amount of nutrients retained by the leaves, which in turn depends on (1) the total leaf area that can be wetted by a sprayed solution and (2) its nutrient concentration. It is therefore unlikely that, before a crop reaches a desired leaf area index (LAI: the ratio between total crop leaf area and soil area), significant quantities of nutrients can be provided to the plant by foliar nutrient application.

Aqueous-based sprays are generally inefficient in wetting and spreading over waxy leaf and fruit surfaces due to high surface tension (at 20°C, surface tension at the air–water interface is $72.8\,\mathrm{mN\,m^{-1}}$). Because foliar nutrients are often included in standard pesticide sprays, they take advantage of the formulation and tank additives included with the pesticide; when they are applied alone, however, surfactants are often added in the tank to reduce surface tension and increase the wetted leaf area. Different volumes of spray solution are applied under field conditions depending on the crop and cultural practicies: for example, in fruit tree plantations, sprayed volumes often vary from 150 to $1500\,\mathrm{l\,ha^{-1}}$. If low volumes are adopted, e.g., to reduce the number of tank fillings needed to spray a given crop acreage, the concentration of solution needed to provide the same amount of nutrient per unit of area is higher than when high spraying volumes are adopted.

When the sprayed solution reaches the leaf, the impacting of spray droplets results either in reflection or retention. When reflected, a droplet may be intercepted by another leaf or fall to the soil. Reflection depends in part on properties of the wetted surface:

while no significant reflection occurs from both abaxial and adaxial leaf surfaces and fruits of peach (*Prunus persica*), the epicuticular wax of plum (*P. domestica*) fruits are known to reflect marked amounts of sprayed solution. As leaf and fruit surfaces vary widely in morphology and fine structure (Figure 1), leaf retention can vary from 5 to 30 μl cm^{-2}. Reducing surface tension of the sprayed solution, while increasing leaf wetting, may reduce retention and increase spray runoff if high spray volumes are applied: the addition of surfactants therefore does not always ensure high efficiency of leaf-nutrient penetration.

Only a fraction of the nutrient retained in the aqueous solution on a leaf's surface after spraying is immediately absorbed by the leaf before drying; the remainder dries out and forms a deposit on the leaf. Penetration also takes place from the deposit, depending on the physiochemical characteristics of the deposit and relative humidity of the microenvironment. The size and distribution of the deposits are important features affecting the nutrient's chances of being taken up later. Nutrient salts have a different point of deliquescence (POD, relative humidity level that causes dissolution of salt residue; Table 2) but are in general hygroscopic, so, once air humidity is higher than POD, the residue will be resolubilized in water and the nutrients again have a chance to penetrate the leaf. Penetration rate, however, declines over time as nutrient concentration in the water solution outside the leaf also declines as a result of uptake. By the same token, the residue is exposed to the risk of runoff if it rains.

Mechanisms of Penetration of Foliar-Applied Nutrients

Nutrient penetration of the leaf surface greatly depends upon leaf structure and chemical composition (Figure 2). The barrier to penetration of foliar-applied substances is mainly represented by the several layers of the cuticle. The outer layer, or cuticle proper, is made up almost entirely of cutin embedded with epicuticular wax, a highly hydrophobic leaf-surface component whose primary function is preventing uncontrolled water losses. Leaf absorption is regulated by the amount, distribution, and chemical composition of such waxes. The presence of cutin, which contains polar groups such as polyesterified hydroxy fatty acids, makes waxes more hydrophilic. The cuticle inner layers are made up of hemicellulose (polyuronides and glycans) pectinaceous substances, proteins, and polysaccharide microfibrils that may form a continuum from the outer surface to the epidermal cell wall, like a conducting channel for polar solutes. Cutin and occasionally waxes are also present in the inner layers. The cuticle is separated from the

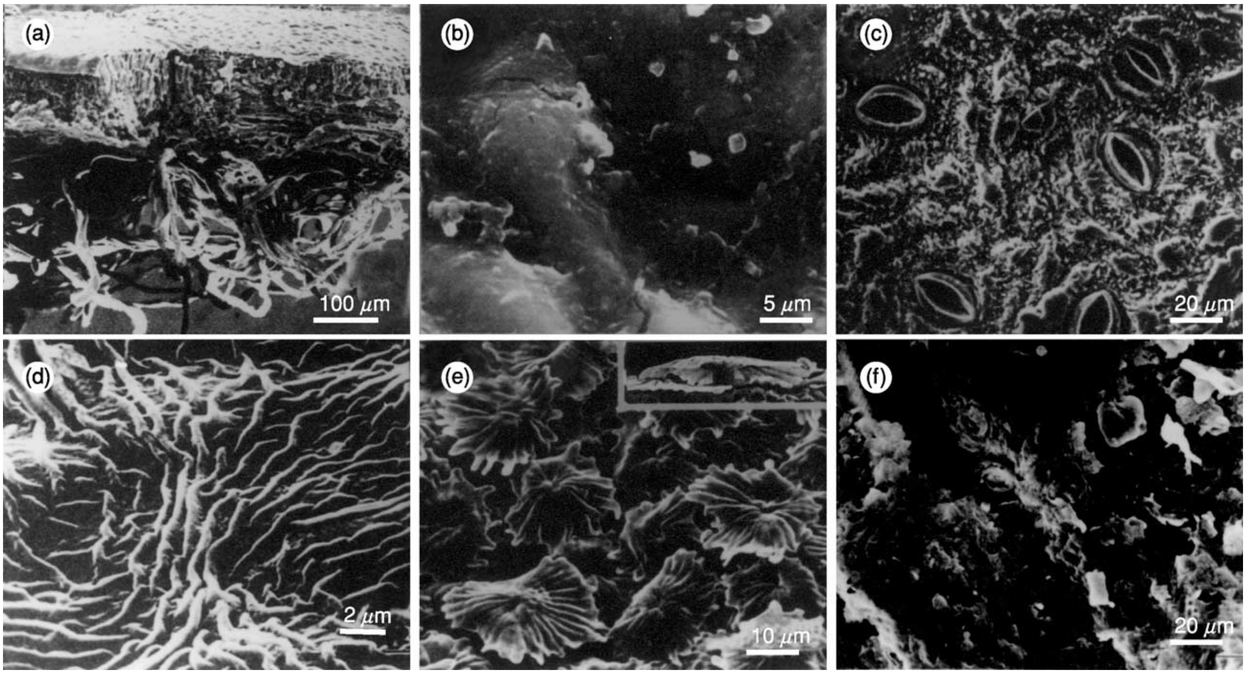


Figure 1 Scanning electron micrographs of the adaxial leaf surfaces of (a) apple and (b) pear; abaxial surfaces of (c) pear, (d) sweet cherry, and (e) olive, with a trichome as inset; and (f) the fruit surface of grapefruit. (Reproduced with permission from Bukovac JM, Cooper JA, Whitmoyer RE, and Brazee RD (2002) Spray application plays a determinant role in performance of systemic compounds applied to foliage of fruit plants. *Acta Horticulturae* 594: 65–75.)

Table 2 Physical properties of selected inorganic salts for foliar nutrition

Salt	*POD (%)*	*Solubility (g kg H_2O^{-1})*	*Molecular weight (g mol^{-1})*
$CaCl_2 \cdot 6H_2O$	33	2790	219
$MgCl_2 \cdot 6H_2O$	33	1670	203
$K_2CO_3 \cdot 2H_2O$	44	1469	174
$Ca(NO_3)_2 \cdot 4H_2O$	56	6600	236
$Mg(NO_3)_2 \cdot 6H_2O$	56	1250	256
NH_4NO_3	63	1183	80
KCl	86	344	75
K_2HPO_4	92	167	174
KH_2PO_4	95	33	136
KNO_3	95	133	101
Calcium propionate $\cdot H_2O$	95	490	204
Calcium lactate $\cdot 5H_2O$	97	31	308
Calcium acetate	100	374	158
$FeCl_3 \cdot 6H_2O$	44	919	270
$Fe(NO_3)_3 \cdot 9H_2O$	54	1500	404
$Mn(NO_3)_2 \cdot 4H_2O$	42	426	251
$MnCl_2 \cdot 4H_2O$	60	1510	180
$ZnNO_3 \cdot 6H_2O$	42	1843	297

POD, point of deliquescence.

Reproduced with permission from Schönherr J (2002) Foliar nutrition using inorganic salts: laws of cuticular penetration. *Acta Horticulturae* 594: 77–84.

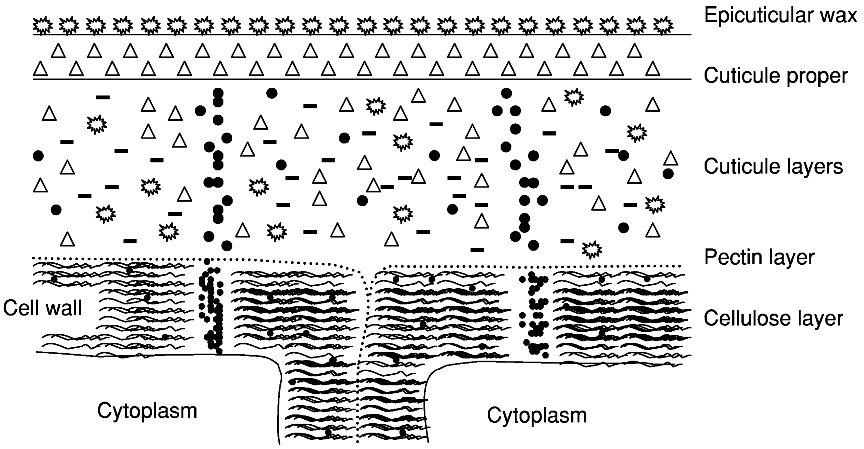


Figure 2 Leaf surface is a mix of wax (jagged symbols), cutin (triangles), pectin (dots), and cellulose (dashed line) arranged in different layers. The hydrophilic fraction (pectin) may form pores that allow a faster penetration of water-soluble compounds.

underlying epidermal cell wall by a pectic layer. Because of its chemical composition, the cell wall does not represent a major barrier to nutrients.

Penetration of substances through the cuticle is a diffusive process. Sprayed water solutions adhering to the cuticles have a concentration higher than the water that saturates the apoplastic space, and this concentration gradient represents the main driving force promoting diffusion into the leaf. The mechanisms of cuticle penetration of inorganic salts differ from those of apolar (nonionic) compounds: most foliar-applied fertilizers are inorganic salts, but nonionic products such as urea and synthetic chelates of metal micronutrients or weak acids such as boric acid are also on the market as foliar fertilizers. The penetration of nonionic compounds is best described as diffusion through the cuticle itself: solute mobility decreases with increasing molar volume and increases with temperature, which enhances the fluidity of amorphous cuticular waxes.

Since inorganic salts are strong electrolytes, when dissolved in water they are highly dissociated and surrounded by water molecules that form hydration shells. Nutrients deriving from inorganic salts are thus highly soluble in water but practically insoluble in lipidic phases such as cutin and cuticular waxes. Unlike nonionic compounds, direct ion diffusion through the cuticle does not occur; hydrated ions require aqueous diffusion paths across the cuticle that are provided by the aqueous pores formed by the hydration of polar groups in the cutin (COOH-, OH-, and ester groups). The radius of aqueous pores varies depending on plant species and external conditions: mean values of less than 1 nm and densities in the range of 10^{10} pores cm^{-2} are reported.

Aqueous pores swell or shrink depending on humidity levels, being more open for penetration of water and solutes when high levels of humidity cause cuticle swelling. The diffusion of ions through the cuticle is inversely related to the radius of the hydrated ions. While most studies on cuticular penetration have been carried out using astomatous cuticular membranes, cuticular penetration from peristomatal regions is particularly efficient due to the higher permeability of cuticle over stomata in comparison with astomatous areas. In addition, stomata guard and accessory cells often present an abundant density of channel pores, a fact explaining, at least in part, the positive correlation between penetration rate and stomata density.

Nutrients may also enter the leaf directly through the stomata, but the relative importance of stomata for total leaf nutrient penetration is still a matter of debate. In many plant species, trichomes also represent areas where the entry of nutrients is facilitated because of their low cutinization. Accumulation of sprayed nutrients underneath fruit lenticels has also been reported, suggesting that lenticels represent a preferential penetration pathway in some fruits.

Factors Affecting Cuticular Penetration

Cuticular penetration of nutrients is mainly affected by three groups of factors: the genotype of the plant, the environment, and the application techniques. Among genetic factors, leaf anatomy and chemical composition of leaf surface, a peculiarity of different plant species, influence droplet reflection. Generally speaking, reflection is positively related to the level of epicuticular waxes and adversely related to the number of hairs, which accounts for the fact that young leaves are usually more efficient in spray retention and in absorbing nutrients than old leaves. In particular, surface microstructures leading to microroughness is the critical factor in reducing reflection. Leaves of species having stomata only on abaxial leaf surfaces show a preferential uptake of nutrients from this surface rather than from adaxial surfaces.

Environmental factors affect leaf cuticle absorbtion of nutrients in different ways. Air humidity improves cuticular penetration of hydrated ions by reducing the hydrophobic properties of the cuticle surface, by causing cuticle swelling, by delaying droplet drying, and by maintaining deposits in hydrated form and/or redissolving the deposit of salts. As salts differ in their POD (Table 2), in windy and dry areas salts with low POD are preferred for foliar application.

Light also increases rates of leaf nutrient uptake, an effect depending on the high permeability of cuticular membranes over the guard cells of the open stomata. High light intensity, however, is not a favorable condition for foliar uptake, as it may increase cuticle thickness and the amount of cuticular waxes, as demonstrated in different species of *Brassica* and *Prunus* and in some cereals; this might also explain the higher specific absorption rate recorded in spring than in summer. While not affecting ion penetration, increasing temperature, within a certain range, promotes the movement of apolar solutes through cuticle, an effect depending on the fluidity of waxes.

The addition of surfactants to the sprayed solution improves leaf wetting by lowering the surface tension, thereby increasing the contact between the liquid and the leaf surface. A liquid having a surface tension lower than 15–20 $mN\,m^{-1}$ penetrates spontaneously through leaf stomata. Surfactants able similarly to reduce surface tension show varying effectiveness, depending on other intrinsic properties: in the case of $CaCl_2$ (the most widely adopted salt for preventing Ca-related physiological disorders in fruits), wetting agents such as alkyl-polyglucosides at a concentration of 0.2 $g\,l^{-1}$ improve the penetration rate by a factor of 12, whereas protein-based surfactants and sodium EDTA (a synthetic chelate) reduce the penetration of $CaCl_2$. In some cases, surfactants may be toxic also, as has been demonstrated for fluorocarbon polymers and octylphenoxy polyethoxyethanol.

For nonionic foliar fertilizers, the rate of penetration is increased by plasticizers (such as tributyl phosphate and diethyl sebactate), which cause an increase in fluidity of amorphous cuticular waxes, an effect similar to that of increasing temperature.

The penetration velocity of ions reaching the leaf surface is usually highest initially then tends to level off over time. Experimental studies carried out with isotopes, which allow a high accuracy of determination, indicate that the half-time of penetration of $^{45}Ca(NO_3)_2$ (6 $g\,l^{-1}$) in pear leaf disks is 18 h, while most salt-derived Ca is absorbed in less than 100 h. In the case of N, urea is taken up faster than other N salts (i.e., KNO_3) because of its apolar nature. The penetration half-time of urea-^{15}N in fruit tree leaves (such as apple, grapefruit, and red raspberry), is approximately 30–50 h, and as much as 90% of the retained urea-N is found in leaves after 5 days. The effectiveness of foliar penetration also depends on salts: for example, apple leaves absorb P according to the following array: $H_3PO_4 > K_2HPO_4 > NaH_2PO_4 > KH_2PO_4 > Ca(H_2PO_4)_2$.

The pH of the sprayed solution affects the rate of penetration depending on plant species and spray solute. A large range of pH (2–10) has been investigated:

Table 3 Compounds often used for foliar fertilization with correspondent application rates

Nutrient	*Compound*	*Nutrient concentration (%)*	*Compound rate (kg ha^{-1})*	*Compound concentration (g l^{-1})*
B	Boric acid (H_3BO_3)	10	3–12	0.6–1.5
	Polyborate ($Na_2B_4O_7$)	20	3–6	0.6–1
Ca	$CaCl_2 \cdot 2H_2O$	27	2–5	2–3.5
	$Ca(NO_3)_2 \cdot 4H_2O$	15.5	4–8	5–6
Fe	$FeSO_4$	37	0.75–2	0.5–1.5
	FeDTPA/EDTA	4–6	3–5	3
Mg	$MgSO_4 \cdot 7H_2O$	9.6	7.5–15	5–12
	$MgCl_2 \cdot 6H_2O$	25	1.5–4.5	1–3
	$Mg(NO_3)_2$	16	4.5–9	3–6
Mn	$MnSO_4H_2O$	32.5	1–3	0.5–2
N	Urea ($CO(NH_2)_2$)	46.6	2.5–15	2–40
	NH_4NO_3	26–27	4	2–4
	KNO_3	13	7.5	5
	$Ca(NO_3)_2 \cdot 4H_2O$	13	7.5	5
P	NH_4PO_4	26.6	0.5–1	0.5
	KH_2PO_4	22.7	1–1.5	0.6
K	KH_2PO_4	29	10	8–10
	KNO_3	38.7	7.5	5–8
Zn	$ZnSO_4$	58	0.3–2	1–2
	$ZnCl_2$	50	0.4–2.4	1–2
	ZnEDTA	2.5	0.6–1	0.6–0.8

DTPA, diethylenetriaminepentaacetic acid; EDTA, ethylenediaminetetraacetic acid.

low pH values are considered optimal for metal ions (Zn), intermediate values (5.4–6.6) for urea, while high values (7–10) are the optimum for potassium phosphate.

The volume of spray solution (and consequently the concentration of the applied nutrients) can affect the rate of cuticular penetration. Generally, higher volume (i.e., 1000–1500 l ha^{-1}) increases the wetting period and delays the time of deposit formation. Low volume may be preferred when surfactants are added to the spray solution; in this case the nutrients retained will have a higher concentration and the deposits will be more uniformly distributed on the leaf. Experimental data for commercial apple trees show a higher uptake rate of urea-N at a concentration of 2 g of urea l^{-1} compared with 20 g l^{-1} and 40 g l^{-1}. A list of selected compounds suitable for foliar application is reported in Table 3.

Plant Use of Foliar-Applied Nutrients

Once nutrients have penetrated the cuticle and reached the apoplast of the external mesophyll cells, they must cross the plasma membrane, because, with some exceptions, it is within the cytosol that they play their role in plant metabolism. Crossing the plasma membrane for nonelectrically charged molecules such as urea is likely to be a simple diffusion, driven by the higher apoplastic concentration. In the cytosol, the urea is quickly hydrolyzed by urease and the resulting ammonium metabolized in the glutamate–oxoglutarate aminotransferase (GOGAT) pathway.

When a mineral ion derived from foliar application reaches the free space of mesophyll cells, it presumably mixes with nutrients already present derived from internal storage remobilization or root uptake. Anions and cations cross the plasma membrane via transporters or via channels. The uptake of a given ion into the cell lowers its concentration in the apoplast and increases the concentration gradient across the cuticle, resulting in higher nutrient penetration.

Some physicochemical characteristics of the apoplast are very important for entering the cytosol: iron, for example, is mainly present in the apoplast in its oxidized state, Fe(III), while it crosses the plasma membrane only after being reduced to Fe(II), a process greatly affected by pH and the organic acid concentration of the apoplast; if the iron is not reduced, it accumulates in the apoplast. The effectiveness of a foliar-applied iron compound is thus highly dependent upon the conditions affecting leaf uptake. Lowering the pH of a spray solution makes the inactivated iron pools in the apoplast available for uptake and may cause regreening of iron-deficient leaves.

The fate of foliar-applied nutrients that are mobile in the phloem is not limited to the leaves, where the absorption process takes place, as other plant organs such as reproductive or storage structures benefit from foliar supply. Nutrients vary as to their phloem mobility, being higher for nitrogen (as amino acid), K, and P than for Fe, Ca, and Mn. Phloem mobility also depends on plant characteristics: for boron, traditionally considered phloem-immobile, a significant phloem mobility is found in plants using sugar alcohols, such as sorbitol in apple, for phloem transport of C. The amount of foliar-absorbed nutrients subsequently exported to other plant organs may differ according to the physiological stage of growth and the presence of major sinks. In perennial deciduous plants, when leaves represent sources of N and storage organs represent major sinks, a significant proportion of N (80–87% in cherry and 50–60% in apple and peach) derived from foliar sprays is transported to the rest of the tree before leaf senescence occurs, and most of this can be recovered in winter in the root system. Nitrogen absorbed by leaves from sprays applied in the spring and summer is exported at varying rates depending on the N demand of newly developing leaves and reproductive organs.

List of Technical Nomenclature

$\mu l\,cm^{-2}$	Leaf retention
degrees Celsius	Temperature
$mN\,m^{-1}$	Surface tension
percentage relative humidity	Point of deliquescence

See also: **Calcium and Magnesium in Soils**; **Fertilizers and Fertilization**; **Iron Nutrition**; **Macronutrients**; **Micronutrients**

Further Reading

Brown P (2001) Transient nutrient deficiencies and their impact on yield – a rationale for foliar fertilizers? *Acta Horticulturae* 564: 217–223.

Bukovac JM, Cooper JA, Whitmoyer RE, and Brazee RD (2002) Spray application plays a determinant role in performance of systemic compounds applied to foliage of fruit plants. *Acta Horticulturae* 594: 65–75.

Eichert T and Burkhardt J (2001) Quantification of stomatal uptake of ionic solutes using a new model system. *Journal of Experimental Botany* 52: 771–781.

Franke W (1967) Mechanism of foliar penetration of solutions. *Annual Reviews of Plant Physiology* 18: 281–300.

Haynes RJ and Goh KM (1977) Review on physiological pathways of foliar absorption. *Scientia Horticulturae* 7: 291–302.

Mengel K (2002) Alternative or complementary role of foliar supply in mineral nutrition. *Acta Horticulturae* 594: 33–47.

Oland K (1960) Nitrogen feeding of apple trees by post-harvest urea sprays. *Nature* 185: 857.

Schönherr J (2002) Foliar nutrition using inorganic salts: laws of cuticular penetration. *Acta Horticulturae* 594: 77–84.

Swietlik D and Faust M (1984) Foliar nutrition of fruit crops. *Horticultural Reviews* 6: 287–355.

Tagliavini M, Millard P, and Quartieri M (1998) Storage of foliar absorbed N and remobilisation for spring growth in young nectarine (*P. persica* var *nectarina*) trees. *Tree Physiology* 18: 203–207.

Toselli M, Thalheimer M, and Tagliavini M (2002) Leaf uptake and tree partitioning of urea-N as affected by concentration and volume of sprayed solution and leaf age of apple trees. *Acta Horticulturae* 594: 591–594.

Weinbaum SA (1988) Foliar nutrition of fruit trees. In: Neumann PM (ed.) *Plant Growth and Leaf-Applied Chemicals*, pp. 81–100. Boca Raton, FL: CRC Press.

Weinbaum SA, Brown PH, and Johnson RS (2002) Application of selected macronutrients in deciduous orchards: physiological and agro technical perspectives. *Acta Horticulturae* 594: 59–64.

FOOD–WEB INTERACTIONS

P C de Ruiter, University of Utrecht, Utrecht,
The Netherlands
J C Moore, University of Northern Colorado,
Greeley, CO, USA

Introduction

Soils harbor a large part of the world's biodiversity. Microorganisms such as bacteria and fungi are by far the most dominant groups of soil organisms, in terms of diversity, numbers, and biomass, but there is also a large variety of micro- and mesofauna such as protozoa (amebae, flagellates, ciliates), nematodes (bacterivores, fungivores, omnivores, herbivores, and predators), microarthropods, e.g., mites (bacterivores, fungivores, predators) and collembola (fungivores and predators), enchytraeids, and earthworms. Trophic interactions among all these soil organisms are known to govern major components in the global cycling of materials, through the decomposition of soil organic matter and the mineralization of nutrients. As these processes determine the availability of nutrients that are limiting to plants, the belowground food web interactions also influence aboveground primary productivity and carbon sequestration. In addition to their roles in shaping the dynamics and persistence of many populations within the soil community, patterns in the strength of the trophic interactions are known to be crucial for the stability of communities and hence the maintenance of biological diversity.

Trophic Structures in Soil Food Webs

Because of the large diversity of soil organisms, soil food-web descriptions are often constructed by adopting the functional group approach. In this approach, species are aggregated into functional groups according to properties that are important to understand their role in ecosystem processes, such as principal food source, mode of feeding, reproduction rate, defenses against predation, or distribution in the soil profile. Such grouping of species into functional units has been supported by analyses of soil food webs, indicating that the behavior of organisms within functional groups seems more similar than that of organisms within taxonomic units, i.e., taxonomy is not necessarily related to function. An example of a functional group description of a soil food web is given in Figure 1, referring to the belowground food web from winter wheat fields in a crop rotation system at the Lovinkhoeve experimental farm in the Netherlands. The soil food-web diagram shows two microbial populations and six general trophic categories of belowground fauna: predators, fungivores, bacterivores, detritivores, herbivores, and omnivores (i.e., consumers of animal prey and other resources), making up five trophic levels: (1) detritus and primary production (roots); (2) primary decomposers and herbivores: bacteria, fungi, phytophagous nematodes; (3) consumers of bacteria and fungi: collembola, fungivorous mites and nematodes, protozoa and bacteriphagous nematodes, and mites; (4) intermediate predators: protozoa, predacious nematodes and predacious collembola; and (5) top predators: predacious mites.

During the last decades of the twentieth century, several large, multidisciplinary research programs provided estimates of the population sizes of functional groups from agricultural soils as well as from native and forest soils (Tables 1 and 2). These data show some interesting differences in food-web structure between agricultural soils and native and forest soils. In most agricultural soils, the population sizes in the soil food webs are organized in the form of trophic biomass pyramids (Table 1), while in the forest soils high abundances may occur at the higher trophic levels (Table 2). Across trophic levels, belowground food webs are organized into compartments of species/functional groups, distinguished on the basis of primary energy source: (1) primary producer (root) channel (I), and (2) the detritus channel (II),

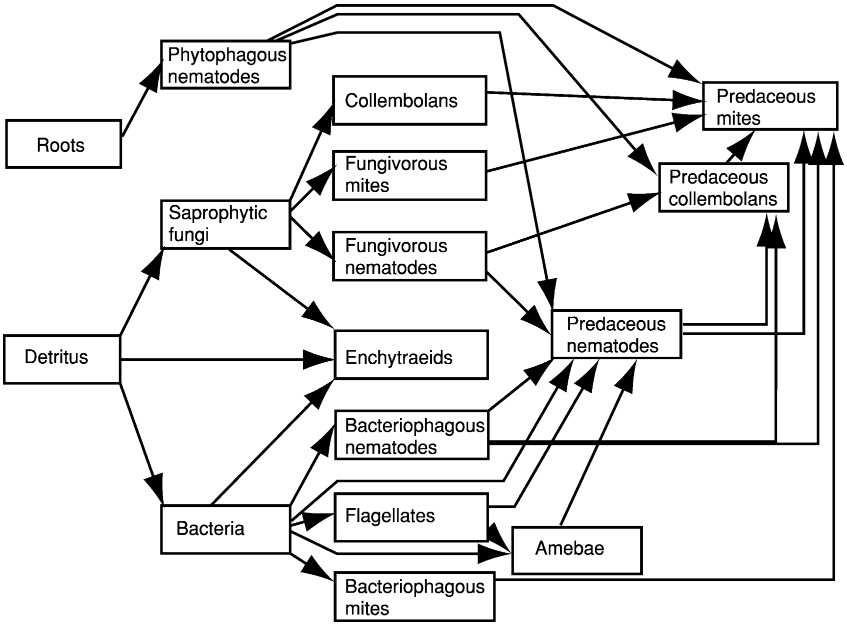


Figure 1 The belowground food web from a winter wheat fields at the Lovinkhoeve experimental farm in the Netherlands. Species are aggregated into functional groups, i.e., based on food choice and life-history parameters. 'Detritus' refers to all dead organic material. Material flows to the detrital pool, though the death rates and the excretion of waste products are not represented in the diagrams, but are taken into account in the material-flow calculations and stability analyses. (Reproduced with permission from De Ruiter PC, Moore JC, Zwart KB *et al.* (1993) Simulation of nitrogen mineralisation in belowground food webs of two winter wheat management practices. *Journal of Applied Ecology* 30: 95–106.)

Table 1 Biomass estimates (kilograms of C per hectare) for the functional groups in agricultural food webs. Values refer to the 0- to 25-cm-depth layer, except for the Horseshoe Bend webs (0–15 cm)

	LH-IF	*LH-CF*	*HSB-NT*	*HSB-CT*	*KS-B0*	*KS-B120*
Microbes	245	228	440	690	740	900
Fungi	3.27	2.12	160	150	1500	2300
Protozoa[a]	20	12	40	50	110	34
Nematodes						
Herbivores	0.4	0.2	0.4	0.5	0.1	0.3
Bacteriovores	0.4	0.3	0.5	1.4	0.5	0.5
Fungivores	0.1	0.1	0.1	0.1	0.2	0.1
Predators	0.1	0.1			0.4	0.4
Arthropods						
Fungivorous collembola	0.38	0.47	0.3	0.1	0.2	0.2
Fungivorous mites	0.05	0.03	1.7	0.5		
Predatory collembola	0.008	0.03				
Predatory mites	0.09	0.06	0.2	0.04	0.2	0.3

[a] Includes amebae and flagellates.

LH, arable land at the Lovinkhoeve (NL); IF, integrated farming; CF, conventional farming; HSB, arable land at the Horseshoe Bend Research Site (Georgia, USA); CT, conventional tillage; NT, no-tillage; KS, Kjettslinge (Sweden); B0, arable land without fertilizer; B120, arable land with fertilizer (120 kg ha^{-1} yr^{-1}).

(Reproduced with permission from De Ruiter PC, Van Veen JA, Moore JC, Brussaard L, and Hunt HW (1993) Calculation of nitrogen mineralisation in soil food webs. *Plant and Soil* 157: 263–273.)

Table 2 Biomass estimates (for CPER and NS, SS, D, and F in kilograms of C per hectare and for NL in micrograms per gram) for the functional groups in natural (prairie and forest) food webs. Values for CPER and NS/SS/D/F refer to the 0- to 10-cm-depth layer

	CPER	*NL-L*	*NL-F*	*NL-H*	*NS*	*SS*	*D*	*F*
Microbes								
Bacteria	304	290	340	264	27	25	19	62
Fungi	63	64	33	14	437	310	385	426
Protozoa[a]	4	130	11	5	11	22	46	30
Nematodes								
Herbivores	3	0.2	0.1	0.02				
Bacteriovores	6	8	1	0.5	0.7	0.14	0.3	0.4
Fungivores	0.4	2	0.2	0.2	0.2	0.07	0.1	0.1
Predators	1	1	0.2	0.1	0.2	0.07	0.2	0.1
Arthropods								
Fungivorous collembola	0.5	73	4	1	2	1	8	7
Fungivorous mites	4	82	18	3	1	2	4	2
Predatory collembola		176	11	1	0.01	0.1	0.1	0.5
Predatory mites	0.4	62	25	7	0.1	0.3	0.7	0.4

[a] Include amebae and flagellates.

CPER, shortgrass prairie; NL, coniferous in Wekerom in the Netherlands; NL-L, litter layer; NL-F, fragmented layer; NL-H, humus layer; NS, SS, D, F, coniferous forest in north Sweden, south Sweden, Germany, and France, respectively.

(Adapted from De Ruiter PC, Van Veen JA, Moore JC, Brussaard L, and Hunt HW (1993) Calculation of nitrogen mineralisation in soil food webs. *Plant and Soil* 157: 263–273; and Schröter D, Wolters V, and de Ruiter PC (2003) C and N mineralisation in the decomposer food webs of a European food transect. *Oikos* 102(2): 294–308.)

where the latter can be split into IIa the bacterial energy channel and IIb the fungal energy channel. In some agricultural soils, the bacterial energy channel dominates, for example the winter wheat fields at the Lovinkhoeve in the Netherlands; in other soils, the fungal channel dominates, for example the barley fields at the Kjettslinge farm in Sweden. Also the forest soils from the Netherlands are bacterial dominated, while the soils from Sweden, Germany, and France are fungal dominated. Both aspects of trophic structure, the pyramid of biomass and the compartmentalization of interactions and energy flows into energy channels, are intimately related to the stability of the system.

Trophic Interactions and Soil Ecosystem Processes

Major soil ecosystem processes such as the decomposition of the soil organic matter and the mineralization of nutrients are the direct result of feeding and processing of material in the soil food web. Because of their dominant numbers (Tables 1 and 2), microbes are usually considered to be the most important contributors to the soil processes, but the faunal groups of organisms are also considered to contribute considerably to the soil processes, despite their relatively low densities. The contribution of the fauna to the cycling of energy, matter, and nutrients depends not only on their own population sizes, or rates of energy use, but also on their ability to influence the functioning of the organisms with which they interact. These influences may occur through a variety of mechanisms. Some faunal groups such as microarthropods break up organic matter by feeding on detritus and by ingesting microbes and adhering detrital material, resulting in an increase in the surface area of the organic matter available for microbial attack. The dissemination of microbes is another potentially important influence of the soil fauna on decomposition processes. Nematodes and microarthropods can passively transport bacteria, fungi, and protozoa in the gut or on the cuticle across regions of soil that are otherwise impenetrable to the microbiota or at greater rates of dispersion than would otherwise be effected. Also in this way, faunal groups can act as catalysts by bringing primary decomposers into contact with substrate. By feeding on microbes, the fauna may affect microbially mediated degradation of the soil organic matter by altering the composition of the microbial community structure through selective grazing, or by affecting microbial growth and metabolic activities by altering nutrient availability. The predatory fauna may also affect ecosystem processes indirectly by decreasing bacterial-feeding nematode populations and subsequently allowing bacterial populations to increase.

Food Web Models

In order to analyze the role of soil food webs and functional groups of soil organisms in soil ecosystem processes, a Detrital Food Web Model has been developed wherein C and N mineralization rates are derived from the feeding rates among the trophic groups using basic elements of the life histories and behaviors of the organisms, and the laws of thermodynamics and conservation of matter. The model was originally developed to analyze N flow rates in the food web from the shortgrass prairie at the Central Plains Experimental Range (Colorado, USA); later it was applied to other soil food webs (see legends to Tables 1 and 2).

The model calculates the feeding rates by assuming equilibrium, i.e., annual growth rates of the populations balance the annual natural death rates and the death rates due to predation:

$$F_j = \frac{d_j B_j + P_j}{a_j p_j} \qquad [1]$$

where F_j is the feeding rate of group j, d_j is the specific death rate of group j, B_j is the annual mean population size of group j, P_j is the death rate of group j due to predation, a_j is the assimilation efficiency of group j, and p_j is the production efficiency of group j. If a predator is considered to feed on more than one prey type, then both the preference of the predator for a given prey and the relative population sizes of the prey types are taken into account:

$$F_{ij} = \frac{w_{ij} B_i}{\sum_{k=1}^{n} w_{kj} B_k} \qquad [2]$$

where F_{ij} is the feeding rate by group j on prey i, w_{ij} is the preference of predator j for prey i relative to other prey types, and n is the number of potential prey types. The model starts calculating the feeding rates of the top predators, because their losses only consist of natural death. Next, the feeding rates of the groups one trophic level below the top predators, whose losses consist of natural death and death due to predation by the predators, are calculated. This procedure is continued working back throughout the web to the lowest trophic levels. C and N mineralization are derived from the feeding rates:

$$C_{ij} = a_j(1 - p_j)F_{ij} \qquad [3]$$

$$N_{ij} = a_j\left(\frac{1}{CN_i} - \frac{p_j}{CN_j}\right)F_{ij} \qquad [4]$$

where C_{ij} is the C mineralization rate by group j feeding on group i, N_{ij} the N mineralization rate by group j feeding on group i, and CN_i and CN_j are the C to N ratios of group i and j, respectively. The model requires data for population sizes, specific death rates, energy conversion efficiencies, preference weighing factors, and C to N ratios. Population sizes are to be established for the particular sites, while most values for the physiological parameters can be derived from the literature, sometimes in combination with adjustments based on the specific environmental conditions. The model provides estimates of the C and N mineralization per trophic interaction, per trophic group and of the community as a whole. In some cases the calculated overall mineralization rates can be compared with observed rates.

The outcome of the calculations in terms of their contribution to N mineralization rates (Tables 3 and 4) show a similar pattern to the population sizes

Table 3 N mineralization rates (kilograms per hectare per year) of the functional groups in the different food webs. Values refer to the 0- to 25-cm-depth layer, except for the Horseshoe Bend webs (0–15 cm)

	LH-IF	*LH-CF*	*HSB-NT*	*HSB-CT*	*KS-B0*	*KS-B120*
Microbes						
Bacteria	52	40	97	137	37	20
Fungi	1	0.8	15	13	19	24
Protozoa[a]	57	35	115	144	54	16
Nematodes						
Herbivores	0.2	0.1	0.1	0.1	0.004	0.06
Bacteriovores	0.9	0.9	0.9	2	2	2
Fungivores	0.08	0.06	0.05	0.03	0.3	0.2
Predators	0.06	0.07			0.3	0.3
Arthropods						
Fungivorous collembola	0.2	0.2	0.1	0.03	0.4	0.4
Fungivorous mites	0.02	0.01	0.6	0.2		
Predatory collembola	0.01	–				
Predatory mites	0.03	0.02	0.08	0.01	0.1	0.2

[a] Include amebae and flagellates.

(Reproduced with permission from De Ruiter PC, Van Veen JA, Moore JC, Brussaard L, and Hunt HW (1993) Calculation of nitrogen mineralisation in soil food webs. *Plant and Soil* 157: 263–273.)

Table 4 N mineralization biomass estimates (for CPER and NS/SS/D/F in kilograms C per hectare per year and for NL in micrograms per gram per year) for the functional groups in natural (prairie and forest) food webs. Values for CPER and NS/SS/D/F refer to the 0- to 10-cm depth layer

	CPER	*NL-L*	*NL-F*	*NL-H*	*NS*	*SS*	*D*	*F*
Microbes								
Bacteria	45	−120	−20	−10	−11	−24	−14	−32
Fungi	8	150	30	5	2	13	46	30
Protozoa[a]	12	300	20	10	10	37	71	61
Nematodes								
Herbivores	1	0.1	0.1	0.1				
Bacteriovores	13	30	3	1	0.3	0.1	0.1	0.2
Fungivores	0.2	2	0.1	0.1	0.01	0.00	0.00	0.00
Predators	0.4	2	0.4	0.1	0.01	0.00	0.00	0.00
Arthropods								
Fungivorous collembola	0.2	0.1	0.1	0.1	0.2	0.2	0.7	0.7
Fungivorous mites	0.8	10	3	0.3	0.06	0.2	0.2	0.3
Predatory collembola	0.003	–	–	–	0.00	0.00	0.03	0.2
Predatory mites	0.1	100	30	10	0.03	0.1	0.2	0.1

[a] Include amebae and flagellates.

(Adapted from De Ruiter PC, Van Veen JA, Moore JC, Brussaard L, and Hunt HW (1993) Calculation of nitrogen mineralisation in soil food webs. *Plant and Soil* 157: 263–273; and Schröter D, Wolters V, and de Ruiter PC (2003) C and N mineralisation in the decomposer food webs of a European forest transect. *Oikos* 102(2): 294–308.)

(Tables 1 and 2). While in agricultural soils, bacteria are usually the most important contributors to N mineralization; in forest soils the model indicates N immobilization by bacteria, due to the low N content of soil organic matter in these soils (Table 4). The calculated N mineralization rates per functional group confirm the expectation that the contribution of faunal groups can be high relative to their contribution to the soil biomass. For example, while microbes may constitute more than 90% of the total soil biomass, the contribution of the soil fauna to N mineralization can exceed that of the microbes (Figure 2).

For some soil food webs, the calculated overall mineralization rates can be compared with observed rates. This verification shows that the model calculates C and N mineralization rates close to the observed rates for the winter wheat fields at the Lovinkhoeve farm and for the barley fields at the Kjettslinge farm (Table 5). For the fields at the Horseshoe Bend research site, the calculated C mineralization rates are close to the observations, but the calculated N mineralization rates are higher than the observed rates. The verification of C and N mineralization rates in the coniferous forest layers in Wekerom in the Netherlands reveals that the

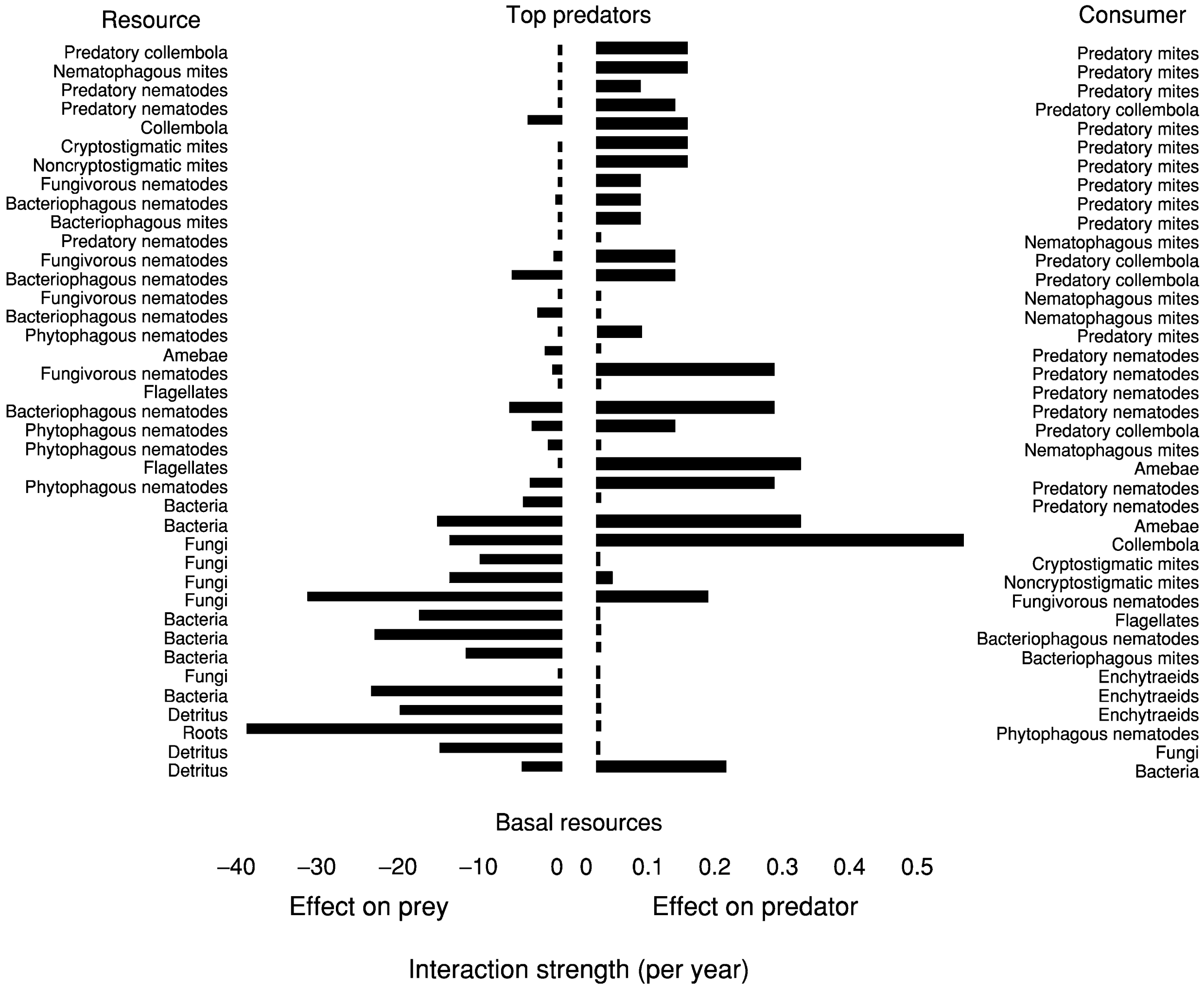


Figure 2 Interaction strengths (per year) along trophic level in the food web from a winter wheat field at the Lovinkhoeve experimental farm in the Netherlands, serving as a representative example. (Reporduced with permission from De Ruiter PC, Neutel AM, and Moore JC (1995) Energetics, patterns of interaction strengths, and stability in real ecosystems. *Science* 269: 1257–1260.)

calculated rates show a similar pattern in differences between the litter, fragmented, and humus layer. For the litter layer, the model underestimates C mineralization by a factor of 2, while for the fragmented and humus layer the model calculates rates close to the observed rates. The calculated N mineralization rates in the litter layer accounts for 85% of the observed rate, while for the fragmented and humus layers this is 50% and 30%, respectively.

Despite the fact that the calculated rates are sometimes close to the observed rates, the model estimates always have to be treated with caution. First, because the model results rely on parameter values that are relatively uncertain, for example the C to N ratio of the substrate for the microbes and the microbial death rates. Second, the indirect effects are only partly taken into account. The model formulations address the stimulatory effects of grazing on microbial growth rates, but other mechanisms behind the indirect effects are at most implicitly taken into account by using observed biomass estimates as input parameters. Despite these weaknesses, the Detrital Food Web Model is regarded as a useful tool in analyzing and understanding the role of the soil food web in soil mineralization processes.

Trophic Interactions and the Dynamics and Stability of Soil Populations and Food Webs

Trophic interactions are likely to affect the distribution and abundance of organisms in fundamental ways, since the success of populations is largely a function of benefits derived from the acquisition of energy (and nutrients) and losses derived from predation. Food web descriptions of the soil community

Table 5 Comparison of the overall C and N mineralization rates (kilogram per hectare per year) as calculated by the food web model with reference values of C and N mineralization published for the different webs

	C mineralization		*N mineralization*	
	Simulated	*Reference value*	*Simulated*	*Reference value*
LH-IF	1890	1690	119	97[a]
LH-CF	1270	1430	78	70[a]
HSB-NT	3970	4800	242	160[b] 90[c]
HSB-CT	4800	5200	300	200[b] 100[c]
KS-B0	1990	1700	120	70[d] 80[e]
KS-B120	1660	1700	70	80[d] 90[e]
CPER	1690	–	87	–
NL-L	25	50	0.7	0.85
NL-F	5	5	0.1	0.2
NL-H	2	2	0.02	0.06
NS	770	–	<1	–
SS	1500	–	30	–
D	2800	–	100	–
F	2600	–	60	–

[a] Obtained from *in situ* measurements.
[b] Obtained from an N budget for 1978–1979.
[c] Obtained from an N budget for 1979–1980.
[d] Obtained from a simulation model.
[e] Obtained from an N budget.

therefore provide a way to analyze the dynamics and persistence of the various populations in the context of the stability of the community as a whole. Central in analyses of the role of trophic interactions in community stability are the interaction strengths. Interaction strengths refer to the per capita – in this case per biomass – effects upon one another. The interaction strengths can be derived from the population sizes and energy flow rates (i.e., the feeding rates; see eqns [1] and [2]) by assuming Lotka-Volterra equations for the dynamics of the functional groups:

$$\dot{X}_i = X_i\left(b_i + \sum_{j=1}^{n} c_{ij}X_j\right) \qquad [5]$$

where X_i and X_j represent the population sizes of group i and j, respectively, b_i is specific rate of increase or decrease of group i, and c_{ij} is the coefficient of interaction between group i and group j. Mathematically, interaction strengths are defined as the entries of the Jacobian community matrix (α_{ij}) being the partial derivatives near equilibrium: $\alpha_{ij} = (\partial\dot{X}_i/\partial X_j)*$. Values for the interaction strengths can be derived from the equilibrium descriptions by equating the death rate of group i due to predation by group j in equilibrium, $c_{ij}X_i^*X_j^*$, to the mean annual feeding rate, F_{ij} (eqn [2]) and the production rate of group j due to feeding on group i, $c_{ji}\ X_j^*X_i^*$, to $a_j\ p_j\ F_{ij}$. With equilibrium population sizes, X_i^*, X_j^*, assumed to be equal to the observed annual mean population sizes, B_i, B_j, the effect of predator j on prey i is:

$$\alpha_{ij} = c_{ij}X_i^* = -\frac{F_{ij}}{B_j} \qquad [6]$$

and the effect of prey i on predator j is:

$$\alpha_{ji} = c_{ji}X_j^* = \frac{a_j p_j F_{ij}}{B_i} \qquad [7]$$

Estimates of the interaction strengths obtained this way for the soil food webs reveal patterns along trophic position, characterized by relatively strong top-down effects at the lower trophic levels and relatively strong bottom-up effects at the higher trophic levels (**Figure 2**). The patterns of interaction strengths are important to the community stability as is indicated by a comparison between the stability of community matrix representations of seven soil food webs (from the prairie and arable soils), using the empirically based values of interaction strengths ('real' matrices) and that of matrices in which these values are randomized. The comparison shows that matrices including the realistic patterns of interaction strengths have a much higher level of stability than their randomized counterparts (**Figure 3**).

The stabilizing patterns of the interaction strengths are the direct result of patterns in the energetic properties of the food webs such as the population sizes

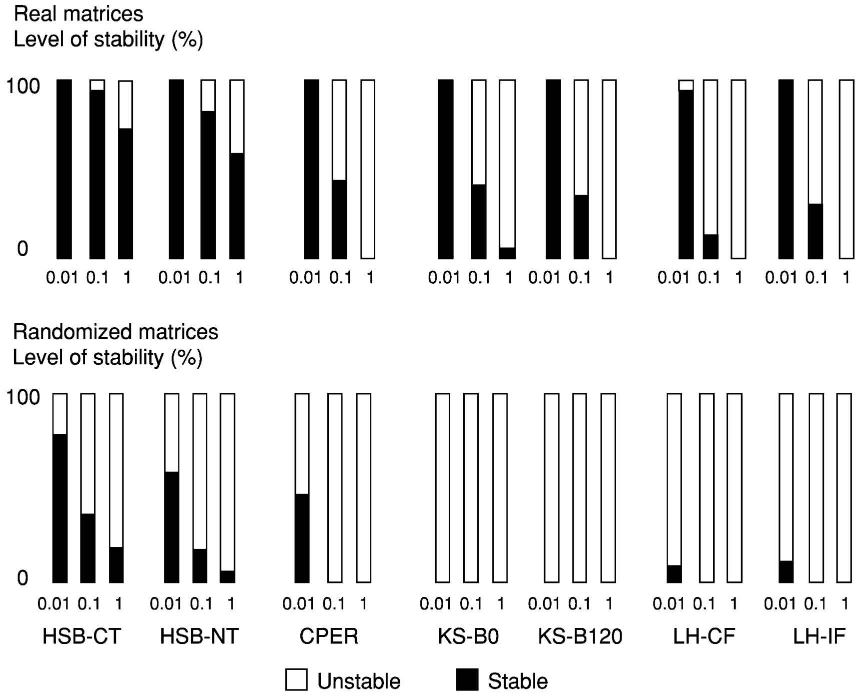


Figure 3 The effects of the patterning of interaction strengths on the stability of seven soil food webs from prairie and arable land. The black fraction in the bars denotes the level of stability (percentage) based on 1000 model runs. Stability of the community matrices is established by evaluating the signs of eigenvalues of the matrices; when all real parts are negative, the matrix is stable and the food web is considered to be locally stable. HSB, Horshoe Bend site, arable land; CT, conventional tillage; NT, no tillage; CPER, shortgrass prairie; KS, Kjettslinge; B0, arable, no fertilizer; B120, arable, with fertilizer; LH, Lovinkhoeve, arable; CF, conventional farming; IF, integrated farming. (Reproduced with permission from De Ruiter PC, Neutel AM, and Moore JC (1995). Energetics, patterns of interaction strengths, and stability in real ecosystems. *Science* 269: 1257–1260.)

(biomasses) and feeding rates (eqns [6] and [7]). Therefore, when the soil biology is looked at in terms of trophic interactions in the soil food web, the structure of the community and the dynamics of the soil populations become inextricably interrelated with soil ecosystem processes and functioning.

Further Reading

Andrén O, Lindberg T, Boström U, Clarholm M, Hansson A-C, Johansson G, Lagerlöf J, Paustian K, Persson J, Petterson R, Schnürer J, Sohlenius B, and Wivstad M (1990) Organic carbon and nitrogen flows. *Ecological Bulletin* 40: 85–125.

Berg MP, De Ruiter PC, Didden WAM, Janssen MPN, Schouten AT, and Verhoef HA (2001) Community food web, decomposition and nitrogen mineralisation in a stratified Scots pine forest soil. *Oikos* 94: 130–142.

Coleman DC, Reid CPP, and Cole CV (1983) Biological strategies of nutrient cycling in soil ecosystems. *Advances in Ecological Research* 13: 1–55.

de Ruiter PC, Van Veen JA, Moore JC, Brussaard L, and Hunt HW (1993) Calculation of nitrogen mineralization in soil food webs. *Plant Soil* 157: 263–273.

de Ruiter PC, Neutel AM, and Moore JC (1994) Modelling food web and nutrient cycling in agro-ecosystems. *Trends in Ecology and Evolution* 9/10: 378–383.

de Ruiter PC, Neutel AM, and Moore JC (1995) Energetics, patterns of interaction strengths, and stability in real ecosystems. *Science* 269: 1257–1260.

Hendrix PF, Parmelee RW, Crossley Jr, DA, Coleman DC, Odum EP, and Groffman PM (1986) Detritus food webs in conventional and no-tillage agroecosystems. *Bioscience* 36: 374–380.

Hunt HW, Coleman DC, Ingham ER, Ingham RE, Elliott ET, Moore JC, Rose SL, Reid CPP, and Morley CR (1987) The detrital food web in a shortgrass prairie. *Biology and Fertility of Soils* 3: 57–68.

Macfayden A (1963) The contribution of the microfauna to total soil metabolism. In: Doeksen J and van der Drift J (eds) *Soil Organisms*, pp. 3–17. Amsterdam: North Holland.

Moore JC and De Ruiter PC (1991) Temporal and spatial heterogeneity of trophic interactions wthin belowground food webs: an analytical approach to understand multidimensional systems. *Agriculture, Ecosystems and Environment* 34: 371–397.

Moore JC and Hunt HW (1988) Resource compartmentation and the stability of real ecosystems. *Nature* 333: 261–263.

Moore JC, Walter DE, and Hunt HW (1988) Arthropod regulation of micro- and mesobiota in belowground detrital food webs. *Annual Review of Entomology* 33: 419–439.

Moore JC, de Ruiter PC, and Hunt HW (1993) Influence of productivity on the stability of real and model ecosystems. *Science* 261: 906–908.

Schröter D, Wolters V, and de Ruiter PC (2003) C and N mineralisation in the decomposer food webs of a European forest transect. *Oikos* 102: 294–308.

Verhoef HA and Brussaard L (1990) Decomposition and nitrogen mineralization in natural and agro-ecosystems: the contribution of soil animals. *Biogeochemistry* 11: 175–211.

FORENSIC APPLICATIONS

W F Rowe, George Washington University, Washington, DC, USA

Introduction

Forensic scientists encounter soil both as trace evidence and as a matrix in which other physical evidence may be found. Trace evidence is the term used by forensic scientists for the minute pieces of physical evidence that can be transferred between the perpetrator and the victim of a crime, between the perpetrator and the crime scene, and between the victim and the crime scene. Trace evidence is used to establish an association among the members of the triad: perpetrator, victim, and crime scene. To give a specific example, suppose traces of soil are found on the clothes of a suspect in a rape case. This soil will be removed in the forensic science laboratory and compared with known soil samples from the crime scene. If the soil from the clothes is indistinguishable from the known sample a positive association between the suspect and the crime scene has been established. Other types of trace evidence include glass, paint, hair, fibers, gunshot residue, explosive residue, ignitable liquids, and biological fluids (such as blood and semen).

The transfer of trace evidence is governed by the Locard Exchange Principle, which states that when two surfaces come into contact there is a mutual exchange of trace evidence between them. Such evidentiary transfers may be primary transfers (i.e., evidence is transferred from surface A to surface B and later recovered from surface B) or secondary transfers, which involve a transfer of evidence from surface A to surface B and from surface B to surface C. Tertiary and even higher-order transfers of trace evidence can occur; however, these transfers present serious interpretive problems for the forensic scientist, because the ultimate source of the trace evidence may be impossible to identify.

Collection of Evidentiary Soil Samples

Known soil samples are collected at crime scenes, as well as any alibi locations. At scenes of crimes, the crime scene investigator collects soil samples from the following locations: (1) the areas immediately surrounding victims' bodies, (2) areas adjacent to any shoe or tire impressions, and (3) any visibly disturbed areas. If the composition of the surficial soil horizons varies within the crime scene areas, representative samples of each soil type are obtained. Because criminal suspects may claim that soil on their clothing or vehicles originated from locations other than the crime scene (e.g., homes or places of work) elimination samples are also obtained from these alibi locations. Clearly the case against a suspect is strengthened if the soil from suggested alibi locations does not match the questioned soil samples. Generally approximately 100 g of soil is sufficient for subsequent laboratory analyses. However, if the soil at a site contains a substantial amount of coarse material, the mass of the known sample is increased so that an adequate sample of the fine fraction is available for subsequent laboratory analyses. Samples are collected from only the top 5–7 cm of the surficial soil horizon. The soil samples are placed in clean containers such as baby-food jars or Ziplok-style plastic bags. The evidence containers are then sealed (with evidence tape or a heat sealer) and marked for identification. Chain-of-custody forms are also initiated for the known soil samples. The locations where the known soil samples have been collected are documented in standard ways: through notes, sketches, and photographs.

The collection of the questioned soil sample is a two-stage process. First, soil from the crime scene is transferred to clothing or vehicles; then the transferred soil is collected at the crime scene, by a crime scene technician, or at the forensic science laboratory. The initial transfer at the crime scene may not produce a representative sample of the soil at that location; for example, clothing rarely retains coarse material such as gravel. Similarly the fines in a shovel load of soil tossed into the open bed of a pickup truck may be blown or washed away before the crime scene investigator has a chance to collect this evidence.

At the forensic science laboratory, trace evidence such as soil is collected in a variety of ways. Hand-picking with a pair of tweezers and a low-power magnifier is used if there is only a small amount of trace evidence, as well as collection using tape lifts or adhesive lifters. Neither of these methods is feasible if an item of clothing has substantial quantities of soil adhering to it, but soil can be scraped gently from an item of clothing on to a sheet of glossy paper. Ideally, scraping should be done in a draft-free, clean room so that air currents do not carry away fine particulates. If an adherent soil layer is thick the laboratory analyst has to consider the possibility that the layer consists of multiple layers deriving from more than one location. Therefore, every effort is made to retain the stratigraphy of the sample and to avoid mixing soil from different layers. A vacuum cleaner with a special filter attachment can also be used to collect soil samples; however, vacuuming is likely to destroy evidence of stratigraphy.

Vehicles require special attention. The crime scene investigators collect soil samples from the vehicle's tires and undercarriage. The interior of the vehicle is also carefully searched, with particular attention being paid to the pedals and floor mats. The vehicle's trunk (boot) is also examined for traces of soil; digging tools with adherent soil may have been placed there.

Laboratory Analysis of Forensic Soil Samples

The methods of soil analysis used in forensic science are predicated on the size of the sample and the use to which the analytical results will be put. The aim of forensic soil analysis is to associate a soil sample taken from an item of clothing or from a vehicle with a specific location. In order to do this, the methods of analysis chosen have to be able to discriminate between soil samples from different locations. Moreover, a method for comparing soil samples has to be practical (i.e., feasible with common laboratory instrumentation), inexpensive, accurate, and applicable to small soil samples. A survey of published materials on forensic soil analysis reveals that this is an area that has yet to be affected by standardization of methodology. Neither the Scientific Working Group on Material Analysis Techniques (SWGMAT) nor American Society for Testing and Materials (ASTM) Committee E30 has promulgated any recommended analytical methods. Consequently, there is considerable variation among forensic science laboratories in their analyses of soil samples, but the methods employed can be placed in three broad categories: (1) methods for comparing soil organic content, (2) methods for comparing particle-size distributions of soil mineral fractions, and (3) mineralogical analysis.

Soil Organic Content

Soil Color

Soil color has been found to be the property of soil that most reflects its pedogenic environment and history. Soil organic matter and iron oxides contribute most to soil color. Organic matter darkens soil, while iron oxides produce a range of soil colors that are dependent on the oxidation state of the iron. The most widely used method for determining soil colors is comparison of soil samples with the color chips in the Munsell soil color charts. In the Munsell color system, color is expressed in terms of hue (basic color), value (lightness or darkness), and chroma (intensity of basic hue). Color determinations are applied to air-dried bulk soil samples; but more elaborate color-determination schemes have also been advocated. Because most of the color of a bulk soil sample is due to the clay fraction, which contains clay particles intimately bound to soil humus (forming the clay–humus complex) and clay particles coated with iron oxides, some forensic soil analysts have suggested basing color determinations on the clay fraction of the soil. Other forensic soil analysts have proposed determining the colors of the soil sample after air-drying, wetting, organic matter removal, iron oxide removal, and ashing. The effectiveness of these treatments in allowing color discrimination of soils from different locations depends on the geographic origins of the soils being compared. A study conducted in the UK by the Home Office Central Research Establishment has found that combining the colors of air-dried samples, moistened samples, and ashed samples results in a high level of discrimination of soils from different locations. On the other hand, a similar study conducted by the Japanese National Research Institute of Police Science has found that color determinations after a different regimen of treatments (air-drying and moistening of bulk samples; organic matter and iron oxide removal from clay fractions) provides the highest discriminatory power. In the latter study, Munsell colors were converted into x-, y-, and z-coordinates: the x-coordinate represents the hue, with basic hues from 2.5YR to 5GY, assigned numerical values from 1 to 10; the *y*-coordinate represents the value; and the *z*-coordinate represents the chroma. Color comparisons between different soil samples are based on the Euclidean distance d^2 between the color coordinates of the soil samples. Pairs of soil samples for which $d^2 \leq 1$ were considered indistinguishable. Using this approach, 97% of 190 pairs of soil samples from different locations were determined to be different.

Percentage Organic Content

The soil organic content can be determined directly by hydrogen peroxide digestion. Properties related to soil organic content can also be examined. Soil pH is commonly measured, although it is known to undergo seasonal changes. Soil polysaccharide content is compared spectrophotometrically. A Fourier transform infrared (FT IR) spectrometric method for the comparison of soil organic fractions has also been developed. In this technique, a potassium bromide disk is prepared from a dried, sieved soil sample. After an FT IR spectrum is obtained, the potassium bromide disk is ground up and placed in a muffle furnace, where the soil organic fraction is burned away. The potassium bromide containing the soil mineral fraction is then reground and pressed into a disk. A second FT IR spectrum is obtained and subtracted from the first to obtain an infrared spectrum of the soil organic fraction.

Soil Particle-Size Distributions

In most forensic science laboratories that conduct soil analyses, particle-size distributions are determined by passing a carefully weighed soil sample through a nest of sieves. Both manual shaking and mechanical shaking are used. Mechanical shakers give more reproducible results than manual shaking; however, experimental parameters such as the torque applied to the stack of sieves in the mechanical shaker and the sample weight can affect the results. In general, obtaining reproducible results with sieving requires careful attention to experimental conditions. A wet sieving method applicable to small soil samples has been developed and is claimed to give more reproducible results than dry sieving.

The UK Home Office Central Research Establishment has conducted a study to develop statistical methods for the comparison of soil samples. Five mock crime scenes were used: at scenes I and II, soil was sampled adjacent to a pair of shoe prints in two gardens; at scenes III and IV, soil was sampled adjacent to pairs of plants in the same two gardens; and, at scene V, soil was sampled adjacent to tire tracks in a rural area. The following analyses were performed: the colors of dry and ashed (850°C for 30 min) soil samples were determined; the pH of each sample was measured; the total polysaccharide content of each sample was determined spectrophotometrically; and particle-size distributions were determined for each sample using both conventional dry sieving and a Coulter counter for the silt fractions. The mock crime scenes were uniform with respect to most measures except total polysaccharide content. The particle-size distributions were compared using an index of variability (IV) and the maximum difference (MD). IV is the difference in the total silt or sand content of two samples, while MD is the maximum difference in the particle-size distributions. From replicate soil samples, the means and standard deviations of each variable were calculated and univariate confidence intervals were determined. Because IV and MD are not independent variables, better confidence intervals were constructed by assuming that IV and MD have a bivariate normal distribution with correlation coefficient ρ. The variables were transformed by a coordinate rotation to produce two normally distributed independent variables. Confidence contours were constructed for the new variables. As a test of this model, 153 pairs of nonduplicate soil samples from the mock crime scenes were compared using a 95% confidence contour; 97% of the pairs were correctly identified as different.

An experiment conducted at the University of Strathclyde, Scotland, UK, illustrates the discriminating power of particle-size measurements when combined with percentage organic matter. A field was divided into 100 cells each measuring 3 m × 3 m. Samples of soil weighing 400 g were collected from the surficial horizon in each cell. The percentage organic content was determined for each of six 1.5-g aliquots of soil from each cell. The dried mineral fraction from each soil aliquot was then sieved. Three variables were selected for comparison: the median value of the particle-size distribution, the percentage of particles in the 90- to 250-μm size range, and the percentage of soil organic matter. Soil samples were selected at random from 20 of the cells and compared with the soil samples from all 100 cells using the *z*-statistic.

Suppose that the soil sample is to be characterized by a variable X having variance σ^2. Measurements of X on m soil samples taken from cell i (one of the c control cells) yield a mean of $\langle x_i \rangle$. Measurements of X on n soil samples from an unknown cell yield a mean of $\langle y \rangle$. The test statistic for the two means being equal is:

$$z_\mathrm{i} = \left| \frac{<x_\mathrm{i}> - <y>}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}} \right| \quad \text{for } i = 1 \ldots c \qquad [1]$$

The most probable cell for the origin of the soil is the control cell for which the probability $P(z_\mathrm{i})$ of obtaining at least the value of z_i is maximum. If the soil has been characterized by N independent variables $X_1 \ldots X_\mathrm{N}$, the most probable cell for the origin of the soil will be that corresponding to:

$$\max\left[\prod_{\mathrm{j=1}} P_\mathrm{j}^\mathrm{N}\,(Z_\mathrm{i})\right] \quad i = 1 \ldots c \qquad [2]$$

Of the 20 randomly selected blind samples, 19 were assigned to the correct control cell using the aggregated P-value. For the one random sample incorrectly assigned, the correct cell had the second-highest aggregated P-value.

The z-statistic method is robust and applicable to many forensic problems. However, the three variables selected in this study may not be the best for soils from another location, and the appropriate variables to use may not be evident *a priori*. Another problem is that statistical significance cannot be assigned to the aggregated P-values.

A study conducted by the Japanese National Research Institute of Police Science has examined the effectiveness of particle-size distributions in distinguishing soil samples from different locations. In this study the soil samples were dry-sieved and the weight percentages of the soil fractions in the following three size categories determined: <0.05 mm, 0.05–0.2 mm, and 0.2–2 mm. Pairs of soil samples were compared by computing the Euclidean distance d^2 between the samples. The threshold value of d^2 to be used to declare paired soil samples to be different was determined from replicate measurements on three of the soil samples. Using this approach, 95.9% of 2628 sample pairs were correctly distinguished as different. When the particle-size distributions were combined with color determinations, 99.5% of the soil samples could be correctly discriminated.

Soil Mineralogical Analysis

In the USA many forensic laboratories have relied on density gradient separations as a useful method for the comparison of the mineral contents of soil samples. Typically, glass tubes are filled with a series of liquids having decreasing densities. The tubes are allowed to stand for several hours so that interdiffusion of the layers can produce a linear density gradient. The samples to be compared are added to the tops of separate density gradient columns and their mineral particles allowed to settle. Mixtures of bromoform and bromobenzene or bromoform and acetone are used to construct the density gradients. The mineral fractions separated in density gradient tubes are recovered and identified by polarized light microscopy or other techniques. X-ray diffractometry has also been used widely for mineralogical analysis. The clay mineral fractions can be isolated from soil samples by sedimentation and their X-ray diffraction patterns can be obtained from oriented mounts. Standard methods for the identification of the clay minerals (dehydration, glycolation, etc.) are used. FT IR can also be used for mineralogical analysis; it is best applied to mineral fractions obtained from density gradient separations.

Soil as a Matrix

A wide variety of evidentiary materials may be found in contact with soil. The bodies of murder victims may be buried; particulate trace evidence such as hairs and textile fibers may be on a buried body or may have been buried seaparately; molecular trace evidence such as ignitable liquids or explosive residues may be dispersed in soil after an arson fire or after an explosion; and deoxyribonucleic acid (DNA) evidence in body fluids and tissues may also have been in contact with soil. The effect of soil on these categories of physical evidence is the subject matter of the newly emerging field of forensic taphonomy.

Buried Remains

The interaction of soil and buried human remains is complex. Decomposing remains contribute to the overall organic content of the soil at the burial site so that the presence of a grave may be inferred by the elevated organic content and darker color of a localized region of a soil horizon. Studies conducted at the University of Tennessee's Anthropological Research Center (also known as the 'Body Farm') have shown that decomposing human remains elevate the temperature of the soil immediately surrounding them and contribute large quantities of aliphatic carboxylic acids to the soil solution. It has been proposed that the relative abundance of certain of these aliphatic acids be used to estimate postmortem interval.

It has been widely recognized for years that burial in soil retards decomposition. This retardation is largely due to the reduction in insect activity. Insects, especially flies, are the main agents for the skeletalization of human remains. While adult flies may feed on carrion, it is mainly fly larvae (maggots) that consume decaying soft tissue. However, even a thin layer of soil may prevent oviposition by most flesh and carrion flies. There are, nevertheless, a few species of flies capable of digging down through soil to deposit eggs on a corpse. Some species of carrion beetle may dig several feet down to reach remains.

The ultimate fate of buried human remains depends on general climatic conditions. Remains in well-drained soils in the temperate climatic zones will eventually be completely skeletalized. If the soil pH is low, even hard tissues such as bone or tooth enamel will ultimately be destroyed; only a soil stain may remain to indicate the presence of a burial. Even bodies in coffins or vaults may be completely destroyed if water infiltrates these enclosures. On the other hand, water can act to preserve buried remains. If the burial site is very damp, adipose tissue in the body may be converted to a light-colored, cheesy mass called adipocere. Adipocere is composed mainly of long-chain aliphatic carboxylic acids. Its formation may be due to the direct hydrolysis of fats by groundwater; however, there exists evidence (primarily from studies of adipocere in Egyptian mummies) that adipocere may also be produced by the action of microorganisms such as members of the genus *Clostridium*. Adipocere may coat the body, discouraging or completely preventing insect attack on the remains. If the soil is very wet and filled with decaying vegetation (e.g., in peat bogs), the soil solution will have a high concentration of tannins and a low oxygen level. Under such anaerobic conditions, skin may be converted to leather and microbial decomposition (putrefaction) of soft tissue may be prevented. 'Bog bodies' may be preserved for hundreds or thousands of years with internal organs more or less intact and available for examination upon autopsy.

The geologic principles of stratification provide a basis for the understanding of buried crime scene evidence. The excavation of graves generally interrupts and mixes soil horizons. Visible evidence of the mixing of the surficial soil horizon with an underlying soil horizon is one of the indicators of the presence of a grave. Graves can also be located using geophysical prospecting techniques such as infrared thermography and ground-penetrating radar. Once a grave has been found, the crime scene investigators have to exercise care during the exhumation of the remains to retain the stratigraphy within the grave. This stratigraphy provides important evidence about the circumstances surrounding the digging of the grave. For example, murderers may spread layers of lime, calcium carbonate, or calcium hydroxide over the remains in the belief that these chemicals will hasten decomposition or reduce the odor of decomposition. The stratigraphy of the burial can be affected by bioturbation, sedimentation, and compression–depression. Scavenging animals may dig down into the infill of the grave to reach the decaying remains; small rodents may also burrow through the grave. Compression–depression of the grave infill leads to slumping of the surface of the grave. The backfill can be tamped down as the perpetrator replaces the soil dug out of the grave. The grave infill may also settle as the soil settles and the remains decompose.

This settling (along with desiccation of the surficial soil horizon after rains) results in cracking of the soil at the edges of the grave. The surface of a grave may also show a secondary depression, caused by an initial, localized expansion, which is the result of the abdomen of the corpse filling with putrefactive gases followed by collapse when the putrefactive gases ultimately escape. If the body is buried facedown a secondary depression may not form. Shallow holes left by scavenging animals may resemble secondary depressions produced by the process of bloating and collapse. Areas of compaction may also be observed within a grave if the victim was buried alive. In struggling to escape the victim may displace soil upwards or sideways. Clearly such evidence would be of the greatest importance in a subsequent murder trial.

When a grave is excavated, investigators have to take care to preserve any impression-type evidence. Shoe impressions may be observed around and within the grave; digging implements such as shovels may leave tool marks on the walls of the grave and on clods of soil removed from the grave; and fabric impressions from the digger's clothing may be left around or within the grave, particularly if the implement used to dig the grave required the digger to kneel while using it.

Trace Evidence in Soil

Contact with soil can alter trace evidence in a number of undesirable ways. The evidence might be completely destroyed or it might be so altered that it can no longer be associated with its original source. The alteration or destruction of trace evidence is caused mainly by soil microorganisms; for example, the restriction endonucleases found in soil microorganisms quickly cut DNA strands into small fragments. Soil microbes can also metabolize the hydrocarbons in ignitable liquids that can be used as arson accelerants. Even if some of the hydrocarbons in the ignitable liquid are detectable in the soil sample, it is sometimes impossible to categorize correctly the type of ignitable liquid used (e.g., light petroleum distillate, gasoline, kerosene, and so forth). In some soils *n*-alkanes are degraded more rapidly than aromatic hydrocarbons, while in others the reverse is true.

The enzyme systems of soil microorganisms can also break down hairs and textile fibers. Among the natural textile fibers, cotton is very rapidly destroyed by contact with soil. As a general rule, in most climates buried cotton fibers are completely destroyed within a few months. Lignified cellulosic fibers (e.g., flax and hemp) are broken down more slowly. The rapid destruction of cellulosic fibers in soil is easily understood: most soils contain microorganisms that use plant debris as sources of energy and of carbon for growth. Buried proteinaceous fibers such as hair, wool, and silk can survive for decades, even in a well-watered soil. Hair and wool fibers, however, can be degraded by soil fungi; fungal hyphae can penetrate hair and wool fibers and grow within the fibers, ultimately completely consuming them. Cellulosic and proteinaceous fibers are more likely to survive if they are buried close to metal artifacts containing copper (e.g., zippers, snaps, button loops, brads, and tacks). Copper ions are well known to have bactericidal effects. Some synthetic fibers also deteriorate rapidly in contact with soil: buried rayons and cellulose acetates are destroyed almost as rapidly as cotton. On the other hand, other synthetic fibers such as nylons and polyesters deteriorate very slowly, presumably because soil microorganisms are unable to elaborate enzymes capable of breaking down the primary polymer structure of these fibers.

See also: **Acidity**; **Archeology in Relation to Soils**; **Bacteria:** Soil; **Clay Minerals**; **Fourier Transform Infrared Spectroscopy**; **Microbial Processes:** Environmental Factors; **Organic Matter:** Principles and Processes; **pH**

Further Reading

Chaperlin K and Howarth PJ (1983) Soil comparison by the density gradient method – a review and evaluation. *Forensic Science International* 23: 161–177.

Dudley RJ (1975) The use of colour in the discrimination between soils. *Journal of the Forensic Science Society* 15: 209–218.

Dudley RJ (1976) The particle size analysis of soils and its use in forensic science – the determination of particle soil size distributions within the silt and sand fractions. *Journal of the Forensic Science Society* 16: 219–229.

Dudley RJ and Smalldon KW (1978) The objective comparison of the particle size distribution in soils with particular reference to the sand faction. *Medicine, Science and the Law* 18: 278–282.

Dudley RJ and Smalldon KW (1978) The evaluation of methods for soil analysis under simulated scenes of crimes conditions. *Forensic Science International* 12: 49–60.

Graves WJ (1979) A mineralogical soil classification technique for the forensic scientist. *Journal of Forensic Science* 24: 323–339.

Hochrein MJ (2002) An autopsy of the grave: recognizing, collecting, and preserving forensic geotaphonomic evidence. In: Haglund WD and Sorg MH (eds) *Advances in Forensic Taphonomy: Method, Theory, and Archaeological Perspectives*, pp. 45–70. Boca Raton, FL: CRC Press.

Murray RC and Tedrow J (1981) *Forensic Geology.* Englewood Cliffs, NJ: Prentice Hall.

Murray RC and Solebello LP (2002) Forensic examination of soil. In: Saferstein R (ed.) *Forensic Science Handbook*, vol. I, 2nd edn, pp. 615–633. Upper Saddle River, NJ: Prentice Hall.

Palenik SJ (1988) Microscopy and microchemistry of physical evidence. In: Saferstein R (ed.) *Forensic Science Handbook*, vol. II, pp. 161–208. Englewood Cliffs, NJ: Prentice Hall.

Sugita R and Marumo Y (1996) Validity of color examination for forensic soil identification. *Forensic Science International* 83: 201–210.

Sugita R and Marumo Y (2001) Screening of soil evidence by combination of simple techniques: validity of particle size distribution. *Forensic Science International* 122: 155–158.

Wanogho S, Gettinby G, Caddy B, and Robertson J (1985) A statistical method for assessing soil comparisons. *Journal of Forensic Science* 30: 864–872.

Wanogho S, Gettinby G, Caddy B, and Robertson J (1987) Some factors affecting soil sieve analysis in forensic science. 1. Dry sieving. *Forensic Science International* 33: 129–137.

Wanogho S, Gettinby G, Caddy B, and Robertson J (1987) Some factors affecting soil sieve analysis in forensic science. 2. Wet sieving. *Forensic Science International* 33: 139–147.

FOREST SOILS

J R Boyle, Oregon State University, Corvallis, OR, USA

Introduction

Forest soils, where soil formation has been influenced by forest vegetation, are generally characterized by deeply rooted trees, significant 'litter layers' or O horizons, recycling of organic matter and nutrients, including wood, and wide varieties of soil-dwelling organisms (Figure 1). There are also soils now covered with forest vegetation, often plantations, on lands that were not naturally forested. These soils are probably undergoing processes that give them 'forest soil-like' characteristics, e.g., litter layers from trees, woody organic residues from deep roots, and associated soil microbe and fauna populations. Like other soils, forest soils have developed, and are developing, from geological parent materials in various topographic positions interacting with climates and organisms. Forest soils may be young, from 'raw' talus, recent glacial till or alluvium, or 'mature,' in relatively stable landscape positions. Just as forest vegetation of the world varies greatly, so do forest soils, e.g., they are shallow, deep, sandy, clayey, wet, arid, frigid, or warm. The following are representative of variability in properties of forest soils, complied for the forest soils of the Douglas fir region of northern California, Oregon (Figure 2), Washington, and southwestern British Columbia: soil depths, 1–2.5 m; nitrogen contained in forest floors, 170–1300 kg ha^{-1}; total N in soils, 1460–22 500 kg ha^{-1}. There are forest soils classified in all the orders of the US soil classification system, reflecting the wide range of conditions in which forests occur.

Forest soils have been studied by many generations of soil scientists. Some studies have been focused mainly on ecologic characteristics, for example on surface organic layers in forests in Denmark, which introduced the terms 'mor' and 'mull'; while other investigations have dealt with nutrients, water supplies, soil organisms (especially mycorrhiza-forming fungi), fertilizer additions, and other impacts of forest management.

Because human land uses have often appropriated the 'best' soils and landscape positions for agriculture, many forest soils are less than optimum in properties that control fertility and potential vegetation productivity. This is illustrated by the designation of many forest soils as 'steep, stony lands' during early soil surveys in the US Lake States. Of course, there are also forest soils with high productive potentials, i.e., with porous and well-aerated root zones, good nutrient-supplying and -retaining capacities, excellent capacities to store plant-available water, and characteristics amenable to robust populations of soil microbes and fauna. And, in many areas of the world today, forests grow on lands once used for

Figure 1 Beech (*Fagus sylvatica*) forest in Germany.

Figure 2 Douglas-fir (*Pseudotsuga menziesii*) forest landscape in Oregon, USA.

Figure 3 *Pinus radiata* plantation in New Zealand. (Photo by B. Dyck.)

intensive agriculture, e.g., parts of northern and central Europe, the northeastern and southeastern USA, and New Zealand (Figure 3).

Many forest soils are shallow and low in organic matter and water-holding capacity, on steep, rocky landscapes, in mountainous areas of the world, e.g., some of the Norwegian and Swedish mountains, the Alps, the Andes, the Sierra Nevada and Rocky Mountains of North America, and in arid environments, e.g., in Mediterranean countries, parts of Chile, Mexico, and the ponderosa pine regions of the western USA. In large areas of Canada, the US Lake States, Norway, Sweden, Finland, and Siberian Russia, forests grow on soils that are shallow, often very wet, on peatlands of several types, and on some soils with permafrost. These forests are slow-growing, due in part to limited nutrient- and water-supplying capacities, shallow rooting potentials, and cold temperatures of the soils.

Roots and Mycorrhizae

Tree roots can penetrate to depths of 10 m or more, depending on the impermeable or porous nature of subsoils and/or bedrock. Deep root systems provide anchorage and stability for trees and provide access to nutrient and water sources far below normal 'soil.' Roots at depth proliferate fine roots and mycorrhizae in thin deposits of soil materials in fissures, providing access to water and nutrient supplies that are not accounted for in evaluations of 'standard' soil profiles. Nutrient uptake from deep roots, transport to and incorporation in plant tissues, subsequent deposition of the tissues as litter, and litter decomposition can contribute to 'upward' cycling of nutrient elements from deep in a soil or the regolith to surface soil horizons.

Mycorrhizae, symbiotic root-fungus structures essential for all forest trees, provide, via their proliferations of fungal hyphae, greatly expanded surfaces for sorption of water and nutrients, and, in some cases, they provide some inhibition of invasions of pathogens. In addition, mycorrhizae facilitate plant uptake of nutrients, especially N and P, from complex compounds not traditionally considered to contain plant-available elements. There is also substantial evidence that mycorrhizae, and other rhizosphere organisms, can facilitate root uptake of nutrients from primary minerals; these processes involve 'dissolution' of mineral structures and nutrient release due to both acidity and complexing characteristics of rhizosphere compounds from fungi, and perhaps from other microbes.

In some forest ecosystems, tree roots represent 30–50% or more of annual primary productivity. Annual or more frequent death and decay of vast proliferations of fine roots in forest soils provide significant additions of organic matter and carbon sequestration, contributing to processes that sustain soil fertility and provide vast stores of carbon. Where there are significant quantities of deep roots, including large, lignaceous roots, these 'ecosystem functions' are difficult to quantify, while being of great significance; decomposition and carbon release processes can be very slow – storage of carbon and nutrients in deep root-derived organic matter is very long-term.

Litterfall and Litter: Organic Matter Recycling in Forest Soils

Forest vegetation adds litter and organic matter to soil surfaces and within soils, because of fauna-mediated incorporation of litter and to the deaths of fine roots and mycorrhizae. These additions are recurrent seasonally and annually in most temperate forests and are continuous and/or episodic throughout the year in many tropical forests. While forest vegetation remains, this litterfall recycles nutrient elements via decomposition processes. In some systems, from 30 to 60% or more of annual tree requirements for nutrients may be met by this recycling. Litter organic matter, interacting with mineral soil particles, and in processes of comminution and decomposition by soil fauna and microbes, contributes to building and maintenance of soil structure and accompanying porosity, and water- and nutrient-holding capacity.

The plant sources, amounts, and compositions of litterfall change during ecosystem succession as forests develop during their regrowth after disturbance of a previous forest or after planting. These variations and dynamics depend on the nature and intensity of site manipulation and management, e.g., elimination of vegetation that competes with desired trees will limit the variety of litter during a forest growing cycle. Because of this repeated and continual litter recycling, whether from complex plant communities or from forest trees in monoculture stands, some forest ecosystems can be self-maintaining for nutrients and soil organic matter for the many years that trees grow and, perhaps, indefinitely. Other forest ecosystems require inputs of some nutrients to sustain desired growth of the trees, e.g., pine plantations on very old, deeply weathered sandy soils of Florida require P fertilizer for satisfactory commercial forest growth.

Accumulated litter layers provide habitats for proliferation of roots and mycorrhizae, and for many organisms, including fauna and microbes that carry out decomposition. Due to the variety in and composition of litter, in any single forest ecosystem decomposition processes are very complex, involve differing groups of organisms, and proceed at different rates (Figure 4). Early stages of decomposition for much litter include comminution – chopping and shredding – of individual plant materials. These processes may be carried out by centipedes, beetles, springtails, worms, slugs, and any of a vast variety of soil fauna, depending on the ecosystem. Of course fungi and bacteria are the ultimate decomposers. In some forest litter layers, nets and mats of fungal hyphae bind together litter components to create carpet-like forest floors. In some tropical forests, litter that reaches the forest floor is already being decomposed and has a very short residence time on the soil surface, creating a rather bare mineral soil surface.

Forest soil scientists recognize three broad categories of litter layers or humus forms: mull, moder, and mor. Mull humus is litter that is rather rapidly comminuted and decomposed, with organic matter being rapidly incorporated into A horizons by soil fauna, with earthworms often playing major roles in decomposition and incorporation processes. Mulls contain what are designated as L and F, or Oi and Oe, horizons. Mor humus types, often forming in coniferous forests with cool, moist – or very dry – forest floor

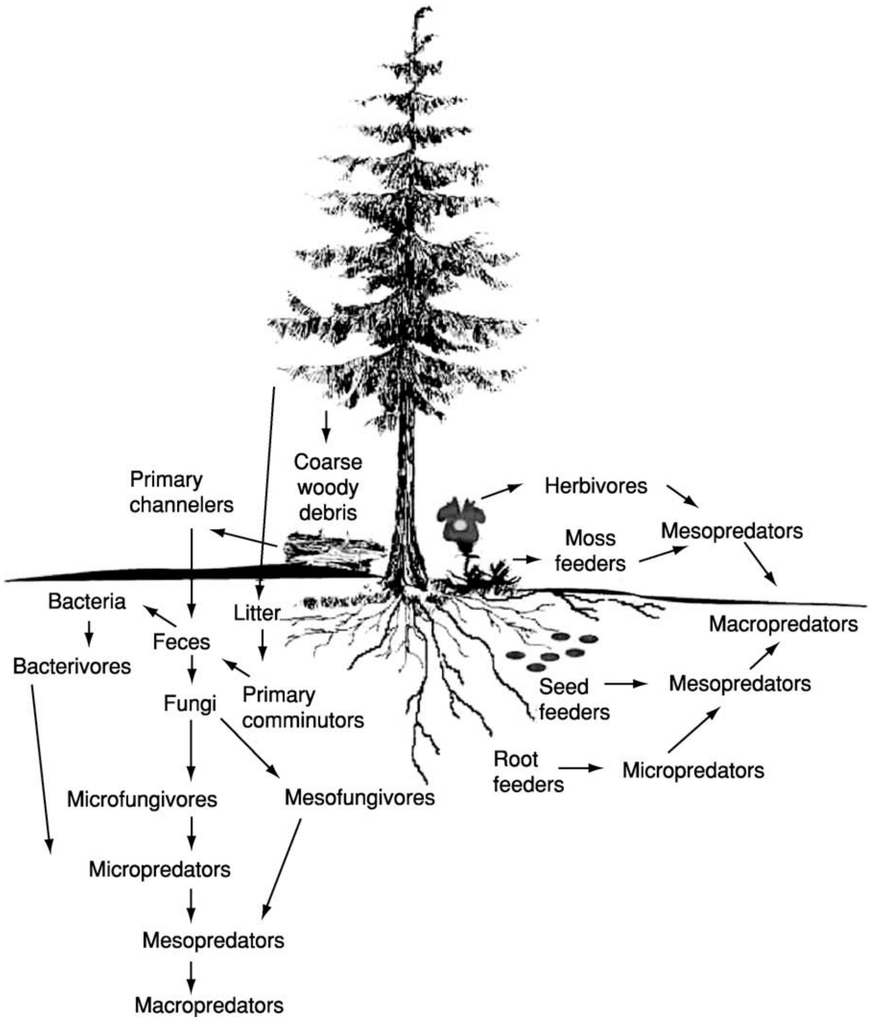


Figure 4 Roles of soil fauna in forest ecosystems. (From a drawing by Andy Moldenke.)

conditions result from litter that is slow to decompose and be incorporated with mineral soil materials; fungi predominate as final decomposers. Mor litter layers, with L, F, and H (Oi, Oe, and Oa) layers and very thin or absent A1 horizons, can accumulate to thicknesses of 15–30 cm or more. Moder humus forms are intermediate in characteristics, with significant accumulations of slowly decomposing materials – L/Oi layers – above F/Oe horizons and an A1 or Ah horizon. Within each of these broad categories of humus forms are numerous variations, depending on the characteristics of the forest ecosystem.

Biogeochemistry and Hydrology

Studies of potential nutrient supplies from forest soils and of nutrient element states and dynamics in forest ecosystems have long been conducted as site-quality classifications, as nutrient-cycling studies, and, since the late 1960s, as research of forest ecosystem biogeochemistry. For many forests, attention has focused on nitrogen, because it is very often the nutrient most limiting for forest growth. Studies of nitrogen biogeochemistry in forests have included: total ecosystem supplies; rates of inputs from precipitation and symbiotic dinitrogen fixation; stores and dynamics in litter and soil organic matter; ammonification and nitrification; sequestration and leaching rates; and denitrification. For some tree species and systems, there has been research of supplies and deficiencies of phosphorus, sulfur, and potassium (e.g., in coarse outwash sands, in glaciated landscapes in the state of New York, in the USA). And, for some systems, researchers have identified limiting quantities of boron (e.g., for *Pinus radiata* in New Zealand) and other micronutrients.

With broad ecosystem and watershed studies, scientists have documented interactions among system hydrology, vegetation, soils, and nutrient elements, especially nitrogen in nitrate form. Removals and suppression of vegetation followed by accelerated decomposition of organic matter, interruption of transpiration flows of water, and alterations of stream water chemistry in forested watersheds have been documented, with significant transfers of nitrate nitrogen to streams in some forest systems, while other systems retain nitrogen rather well after disturbances. These research results are very relevant for sustainable management of forest ecosystems and interacting aquatic systems, and are important components of scientific knowledge of forest–soils interactions and implications for management.

Within the hydrologic cycle and watershed context of forest lands, forest soils contribute to moderation of flows of water, e.g., influencing intake and percolation, limiting runoff, providing filtration, storage, and transmission and delayed flows to streams. Because of litter layers, high porosities associated with soil fauna activities, root proliferation and depth, and many macropores, forest soils have very high infiltration and percolation rates, rates that far exceed rainfall intensities in many ecosystems. These and other attributes of forest ecosystems provide valuable watershed values for the millions of people dependent on forested landscapes for water supplies. The scientific discipline of forest hydrology encompasses these and other aspects of forest–soil–water interactions.

Disturbance Effects on Forest Soils

Because ‘crop rotations’ for forests are of such long duration, e.g., decades to more than 100 years, it is important to consider effects of both natural and man-created disturbances to forest soils when contrasting forest soils with agricultural soils. Disturbances in forest ecosystems influence forest soils and associated dynamics of water, nutrients, and vegetation. Disturbances include windthrow (or blowdown), fires, landslides or other mass movements, and disease or insect outbreaks that defoliate and/or kill many trees. Winds that blow down trees or patches of forests (in some cases hundreds to thousands of hectares) often result in excavation, the tipping up of large root masses with accompanying soil disturbances. This mixes soil horizons, exposes fresh materials to weathering processes, and often creates pit-and-mound microrelief, adding to soil spatial variability, creating numerous microhabitats for subsequent vegetation and contributing to challenges for researchers sampling soils.

Forest fires, with their extreme ranges of intensities, durations, and extents, alter soils indirectly, via depositions of ash, charcoal, and partially burned litter, and directly, depending on temperatures imparted, in effects on soil organisms, root systems, organic matter, chemical characteristics, and physical properties. Losses of vegetation and forest floor organic matter alter system biogeochemistry and hydrology. Nutrient elements, especially nitrogen, may be volatilized in fires and lost from the forest. Residual ash materials provide a pulse of nutrients that may stimulate plant regrowth, or ash may be carried off-site by postfire winds or overland flows of water. With reduced interception, storage, and transpiration capacities owing to removal of vegetation and forest floors, infiltration, percolation, storage, and runoff of water will change; effects vary greatly depending on ecosystem and fire characteristics and postfire weather. Plant succession following a fire can result in rapid uptake and storage of nutrient elements, will contribute to rebuilding forest floors and soil organic matter, and often includes nitrogen-fixing plants that can replenish, to some extent, nitrogen loss during a fire. Because of the complex and variable interactions among fire characteristics and ecosystems, very few generalizations about fire–soil interactions are valid for all forest soils, presenting significant challenges for managers of postfire landscapes, e.g., to replant or not; is erosion protection or nutrient replenishment needed?; what vegetation will result from successional processes?

Diseases in forests (e.g., root rots; needle cast fungi) damage and kill trees, including root systems, contributing dead organic matter on soil surfaces and within soils, and influencing all related ecosystem dynamics of nutrient biogeochemistry, soil structure, and hydrology. Similarly, insect epidemics alter forests and influence soils, e.g., those of defoliating insects (e.g., gypsy moths, Douglas-fir tussock moths) may cause significant, unusual depositions of foliar organic matter to forest floors, with attendant nutrient dynamics and other system effects; large-area tree death due to bark beetles leaves ‘standing dead’ trees that are fire hazards as well as potential soil organic matter.

Human Effects on Forest Soils

Human activities of global, regional, watershed, and site-specific scales influence properties and processes of forest soils. These influences have been well researched by forest-soil scientists, ecologists, chemists, aquatic biologists, and forest engineers, and are only listed here rather than discussed in any detail. Effects of acid rain and related pollutants, including in some cases heavy metals, deposited in forests from the atmosphere have been well documented. Effects include changes in soil acidity, interactions with aluminum chemistry, leaching of cation nutrients, alterations of nitrogen status, influences on root growth, transfers of chemicals to waterways, and effects on soil and aquatic biota.

Forest-management activities include use of heavy machines in forest harvests, site preparation, removal of vegetation and alteration of plant communities, and uses of fertilizers, herbicides, and pesticides. Physical site manipulations such as mechanical removal of competing vegetation and forest floor organic matter, burning, and scarification can alter soil physical properties such as porosity, infiltration capacity, percolation rates, and susceptibilities to erosion. Burning, to eliminate or reduce competing vegetation and forest floor and litter horizons for preplanting site preparation, affects soil properties and processes, as do draining and bedding to alter water tables in relation to rooting zones. Fertilizer additions and additions of herbicides and pesticides affect soil chemistry, organic matter, and biota. These management manipulations can influence forest floors and litter horizons, affect soil porosity, penetrability for roots, hydrology, gas exchanges, chemical properties, organic matter accumulations and dynamics, and soil biota, including potential symbionts of nitrogen-fixing plants.

Forest plantations of single tree species have very different ecosystems compared to many native forests, and create intensive and usually simplifying interactions with soils. These interactions may or may not influence soil properties essential for sustained productivity, e.g., nutrient element biogeochemistry, organic matter dynamics, soil biota, and dynamics of soil structure and porosity. These influences can result when vegetation is extremely simplified by use of site manipulations and herbicides to eliminate plants that would compete for site resources with the desired trees. In the most simplified plantations, with a single tree species as the sole source of organic matter recycled to soils, litter horizons and soil biota will be simplified, resulting in less-complex and -resilient nutrient-cycling dynamics. It is important that forest managers understand the implications of site manipulations for soil properties and processes essential for long-term productivity. With appropriate management, forest soils in such systems can be maintained or enhanced in their productive capacities.

‘Ecosystem Services’ of Forest Soils

In the context of the concept of ‘ecosystem services,’ i.e., beneficial interactions of forest ecosystems with the biosphere, forest soils provide several properties and functions. Among these are the vast reservoirs of stored carbon and the myriad habitats for macro-, meso-, and microorganisms contributing to biodiversity. Watershed and hydrologic functions of forest-soil ecosystems include stabilization of steep slopes and moderation of water flows to streams. Forest managers must be cognizant of these benefits and of the effects of management on altering critical soil properties and processes that influence these and other ecosystem services.

Criteria and Indicators for Soils: Sustainable Forest Management

In the context of criteria for sustainable world ecosystems and societies, the Montreal Process ‘Criteria and Indicators’ for forests and forestry include protection against accelerated soil erosion and maintenance of soil organic matter, as well as attendant soil processes and properties. Soil scientists have developed a variety of indicators of soil quality, i.e., measurements or estimates of soil characteristics considered essential for sustaining forests. These recognize the important functions of soils in influencing carbon and energy fixation in forest growth, nutrient cycling, hydrology, water quality, and biodiversity. Quantification of soil quality or its changes due to management is difficult because of complex, variable processes. The challenge is to identify relatively simple, outcome-oriented indicators for soils, e.g., bulk density, aeration porosity, and organic matter content. In addition to development and testing of indicators, there is a need for fine-scale assessments of soil properties linked to watershed-scale areas, because local, site-specific changes in aggregate over a large area have off-site and cumulative effects. It is also important to stratify land areas based on potential risks to soils imposed by management practices; this can provide a framework for considering management alternatives. Soil-quality indicators can serve as one basis for adaptive forest management designed to meet sustainability objectives for soils, water, and ecosystem productivity.

Summary: Soil Conditions for Forest Productivity

For forests, including aquatic ecosystems, soils provide numerous essential conditions: water-holding, -transmitting, and -supplying capacity; porosity and aeration; physical tilth for root growth; nutrient-supplying and -holding capacities; habitats for symbiotic microorganisms and for myriad soil organisms essential for decomposition and maintenance of structure. Perennial presence of forest vegetation and recycling of organic matter via litterfall and litter decomposition are essential components for productive, sustainable forest ecosystems (Figure 5). A simplified

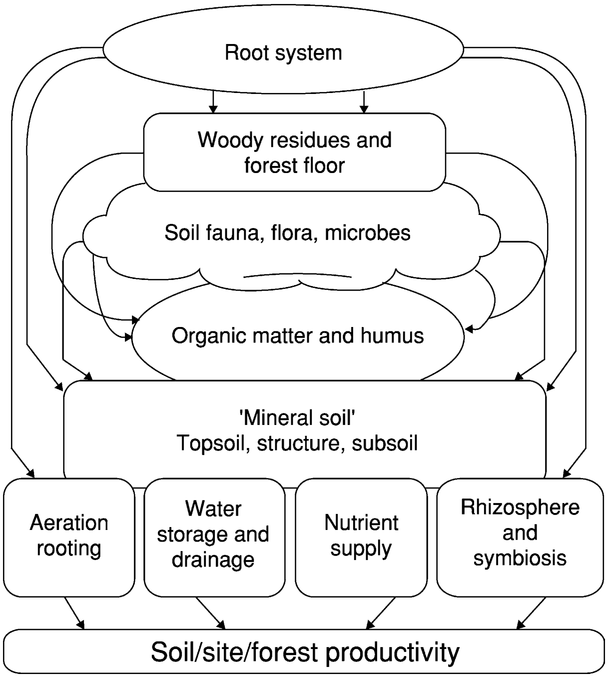


Figure 5 Forest productivity – a view from the bottom.

set of principles for managing forest soils includes these: maintain surface and incorporated organic matter; maintain topsoil – prevent erosion; maintain soil structure for water, air, nutrients, and organisms; and maintain significant land areas of forest-soil ecosystems for all the benefits they provide for the biosphere and human societies.

See also: **Carbon Cycle in Soils:** Dynamics and Management; **Classification of Soils**; **Factors of Soil Formation:** Biota; **Fauna**; **Humification**; **Mycorrhizal Fungi**; **Organic Matter:** Genesis and Formation; **Organic Residues, Decomposition**; **Productivity**; **Sustainable Soil and Land Management**; **Watershed Management**

Further Reading

Attiwell PM and Leeper GW (1987) *Forest Soils and Nutrient Cycling*. Melbourne, Australia: Melbourne University Press.

Boyle JR and Powers RF (eds) (2001) *Forest Soils and Ecosystem Sustainability.* Papers from the Ninth North American Forest Soils Conference, 1998, Tahoe City, Calif. Amsterdam, The Netherlands: Elsevier.

Burger JA and Kelting DL (1998) Soil quality monitoring. In: *The Contribution of Soil Science to the Development of and Implementation of Criteria and Indicators of Sustainable Forest Management*, pp. 17–52. SSSA Special Publication No. 53. Madison, WI: Soil Science Society of America.

Fisher RF and Binkley D (2000) *Ecology and Management of Forest Soils*, 3rd edn. Chichester, UK: John Wiley.

Green RN, Trowbridge RL, and Klinka K (1993) *Towards a Taxonomic Classification of Humus Forms*. Forest Science Monograph 29. Bethesda, MD: Society of American Foresters.

Heilman PE, Anderson HW, and Baumgartner DM (1979) *Forest Soils of the Douglas-Fir Region*. Pullman, WA: Washington State University Cooperative Extension Service.

Lutz HJ and Chandler RF Jr. (1946) *Forest Soils*. New York: John Wiley.

McFee WW and Kelly JM (eds) (1995) *Carbon Forms and Functions in Forest Soils*. Papers from the Eighth North American Forest Soils Conference, 1993, Gainesville, Fla. Madison, WI: Soil Science Society of America.

Powers RF, Tiarks AE, and Boyle JR (1998) Assessing soil quality: practicable standards for sustainable forest productivity in the United States. In: *The Contribution of Soil Science to the Development of and Implementation of Criteria and Indicators of Sustainable Forest Management*, pp. 53–80. SSSA Special Publication No. 53. Madison, WI: Soil Science Society of America.

Ramann E (1893) *Forstliche Bodenkunde und Standortslehre*. Berlin, Germany: Springer-Verlag.

Schoenholtz S, Van Miegoret H, and Burger JA (2000) A review of chemical and physical properties as indicators of forest soil quality: challenges and opportunities. *Forest Ecology and Management* 138: 335–356.

Switzer GL and Nelson LE (1972) Nutrient accumulation and cycling in loblolly pine (*Pinus taeda* L.) plantation ecosystems: the first 20 years. *Soil Science Society of America Proceedings* 36: 143–147.

Tamm O (1950) *Northern Coniferous Forest Soils*. (Translated by Anderson ML.) Oxford, UK: Scrivener Press.

Weetman G and Webber B (1972) The influence of wood harvesting on the nutrient status of two spruce stands. *Canadian Journal of Forest Research* 2: 351–369.

Youngberg CT (ed.) (1965) *Forest-soil Relationships in North America*. Proceedings of the Second North American Forest Soils Conference, Oregon State University, 1963. Corvallis, OR: Oregon State University Press.

FOURIER TRANSFORM INFRARED SPECTROSCOPY

D Peak, University of Saskatchewan, Saskatoon, SK, Canada

Introduction

Fourier transform infrared (FTIR) spectroscopy is a form of vibrational spectroscopy that is useful in the study of a variety of soil chemical processes. In the mid-infrared (mid-IR) range, vibrations arise from many environmentally important molecules such as organic acids, soil organic matter, mineral phases, and oxyanions. It is possible to utilize FTIR spectroscopy as a quantitative analytical method and also as a tool to determine bonding mechanisms in solids and on surfaces. Molecular vibrations can be related directly to the symmetry of molecules, and so it is often possible to determine precisely how a molecule is bonding on surfaces or as a component in a solid phase from its infrared spectrum. Many experimental methods exist for probing samples of various states and at different spectral regions.

The Nature of Molecular Vibrations

The infrared region of the electromagnetic spectrum lies between 10 and 10 000 cm^{-1}, which can be expressed as a frequency of 10^{12}–10^{14} Hz. This frequency overlaps with the frequency of molecular vibrations. When infrared radiation is the same frequency as a molecular vibration, it can induce a transition to a higher energy state. This transition will only occur if it induces a dipole. A true molecular vibration is defined as a movement of the atoms of a molecule that do not change the center of mass and that involve no rotation. Vibrations that obey this definition are known as the 'normal modes.' For a nonlinear molecule, the normal modes may be due to symmetric or asymmetric stretching, and in-plane or out-of-plane bending. Not all of a molecule's normal-mode vibrations will be active when subjected to infrared energy. Instead, the symmetry of a molecule will affect the position and number of infrared-active peaks.

Infrared Peaks of Interest

In the mid-IR range (400–4000 cm^{-1}), there are several extremely important and useful classes of bonds that can be studied. FTIR spectroscopy has been used extensively by both geologists and soil scientists to study clay minerals and other inorganic soil components. The structure of minerals can substantially affect their infrared spectra, and entire textbooks are available with tables of peak positions and figures containing spectra of reference minerals. Water has an intense infrared absorbance in the mid-IR region, which typically is considered a limitation of the technique for environmental research. However, the strong absorbance of water bands has been exploited by researchers who investigated the effects of humidity changes on self-supported films of clay minerals.

Many organic compounds also have diagnostic peaks in the infrared region, and FTIR has been used to study the nature of soil organic matter as well as the adsorption of organic acids. Many metals and metalloids form negatively charged oxyanion molecules in aqueous solution, and FTIR spectroscopy has extensively been used to study their adsorption and precipitation reactions. Figure 1 contains infrared spectra from selected experiments using FTIR to study reactions of: (1) natural organic matter (NOM), and (2) inorganic minerals in soils. Infrared absorbance bands in Figure 1a have been labeled to show how changes in functional-group makeup can give rise to new infrared peaks. Also notice how many of the peaks in the NOM disappear or shift in position when NOM is adsorbed on to a mineral surface. This is due to chemical bonding with carboxylic and phenolic NOM groups and the iron oxide surface. In Figure 1b, a mineral transformation is monitored using FTIR spectroscopy. The initial γ-$\mathrm{Al_2O_3}$ solid phase is not stable in water, and over the 2.5-month experimental period it transforms into a crystalline gibbsite $\mathrm{Al(OH)_3}$ phase.

Infrared Spectral Analysis

A detailed analysis of the theoretical infrared spectrum from first principles is not always required. It is possible simply to compare unknown samples to those of spectra with known molecular configurations to determine the structure of an unknown. This is referred to as spectral fingerprinting. This technique relies on one's ability to synthesize reliably reference compounds of known chemical structure or obtain their spectrum from a library. This approach is reasonably straightforward in the case of organic molecules, as large commercial software spectral libraries are available. In the case of unknown inorganic spectral identification, however, suitable standards can often be difficult to synthesize, and aqueous

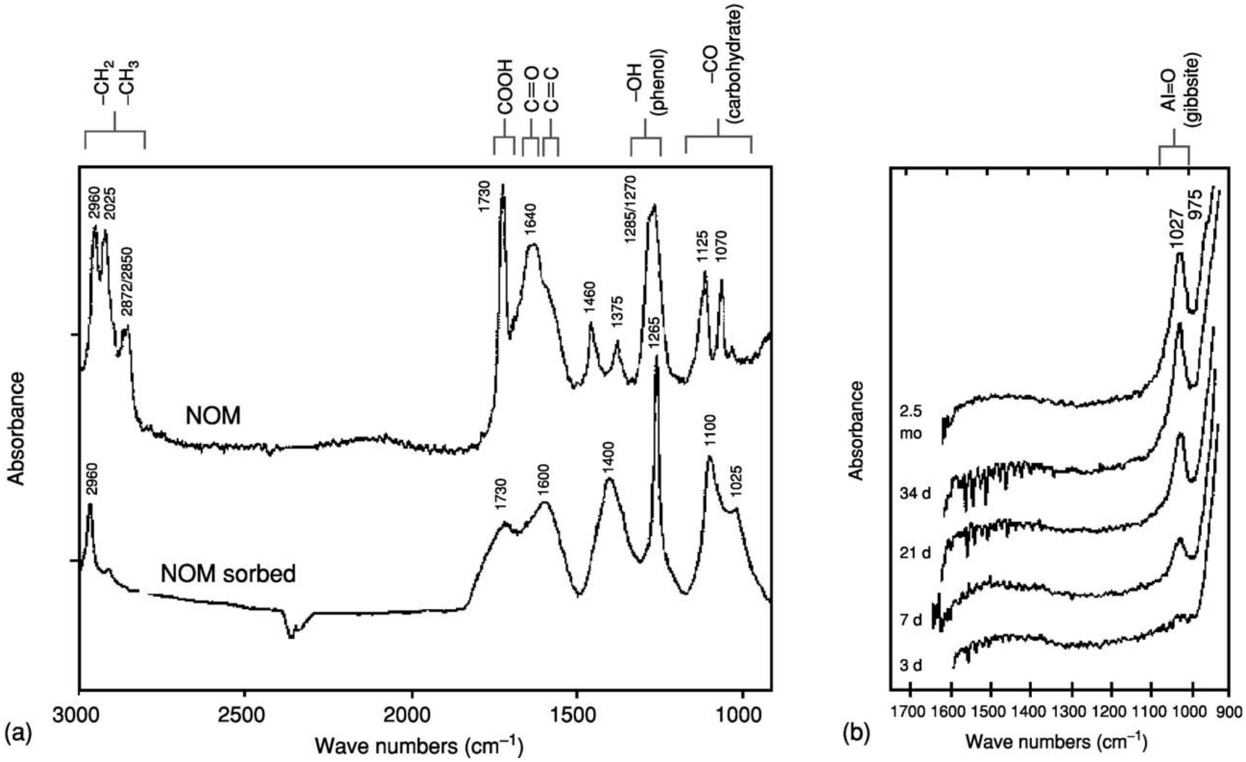


Figure 1 (a) Infrared spectra of natural organic matter in water compared with natural organic matter (NOM) adsorbed on an iron oxide. Reproduced with permission from Gu B, Schmitt J, Chen J, Liang L, and McCarthy JF (1994) Adsorption and desorption of natural organic matter on iron oxide: mechanisms and models. *Environmental Science and Technology* 28: 28–46. (b) Infrared spectra of γ-Al_2O_3 aged from 3 to 75 days. The peak at 1027 cm^{-1} is due to transformation of the initial Al_2O_3 into a crystalline gibbsite phase. Adapted from Wijnja H and Schulthess CP (1999) ATR-FTIR and DRIFT spectroscopy of carbonate species at the aged γ-Al_2O_3 – water interface. *Spectrochimica Acta. Part A Molecular and Biomolecular Spectroscopy* **55**(4): 861–872, with permission.

standards are often present as mixtures of species. For these reasons, assigning peaks based on rules from molecular symmetry is often necessary.

Symmetry rules assign a molecule to a unique point group based upon the arrangement of its atoms in space. A common example of how infrared spectra may be related to molecular symmetry is the case of sulfate. FTIR spectroscopy has been used by many researchers to study sulfate bonding in mineral structures and as surface adsorption complexes. The clear relationship between the molecular symmetry of sulfate species, its point group, and its infrared spectrum is an example of the utility of this approach. In the case of aqueous sulfate, there are two normal modes that are accessible to spectroscopic investigation. They are the nondegenerate symmetric stretching v_1, and the triply degenerate asymmetric stretching v_3 bands. As a free anion in solution, sulfate has tetrahedral symmetry and belongs to the point group T_d. For this point group, only one broad peak, due to the v_3 stretch, is observed at approximately 1100 cm^{-1}. If sulfate is covalently bonded to one or more cations, the symmetry is lowered, a v_1 peak becomes infrared active, and the v_3 band splits into more than one peak. When sulfate is bound as a monodentate complex, the v_3 band splits into two peaks, while the v_1 band becomes fully active at approximately 975 cm^{-1}, and C_{3v} symmetry results. If sulfate forms a bidentate binuclear (bridging) complex, the symmetry is further lowered to C_{2v}, and the v_3 band splits into three bands between 1050 and 1250 cm^{-1}, while the v_1 band is usually shifted to a higher wave number. The relationship between molecular symmetry and the number and position of infrared peaks are detailed in Figure 2. Sulfate is typical of tetrahedral oxyanions, and so the relationship also may be readily extended to other tetrahedral oxyanions of environmental relevance (such as chromate, selenate, phosphate, and arsenate).

Similar rules may be derived for molecules of other point groups also, making the symmetry approach very powerful in the assignment of molecular configuration from FTIR spectra. However, one must be aware that infrared peaks may arise from many potentially occurring molecular vibrations. Of particular concern in environmental science is the fact that interactions with protons will substantially affect the infrared spectrum of sulfate, not only when

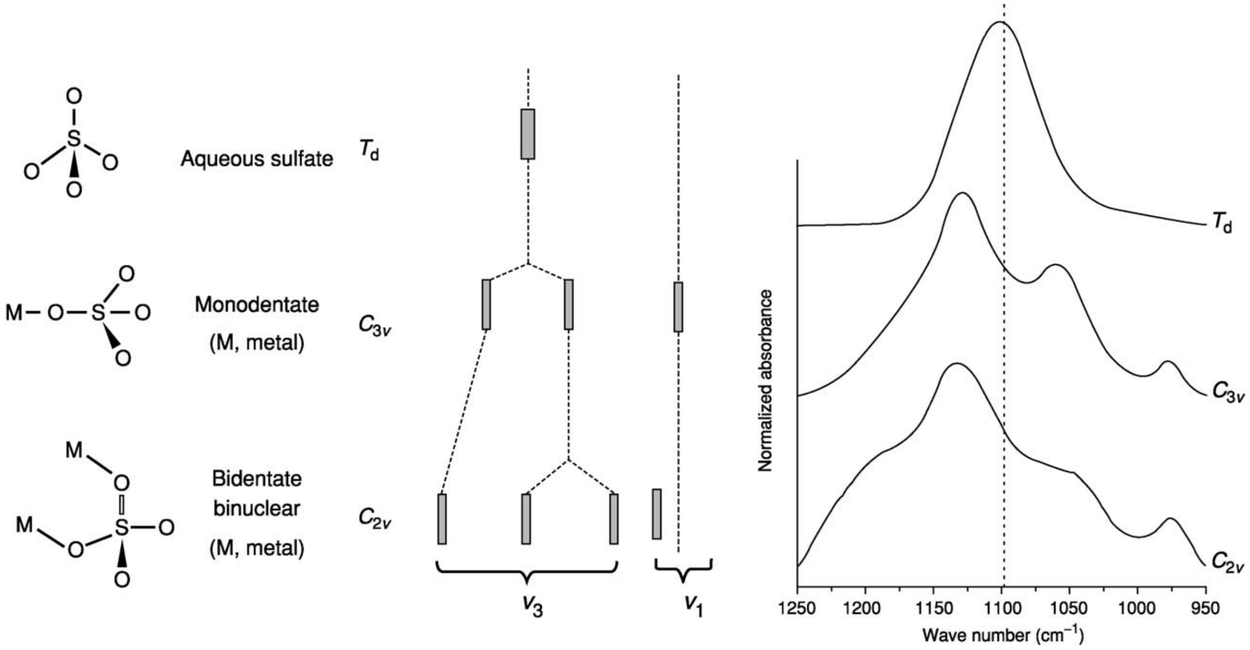


Figure 2 Relationship between molecular symmetry and infrared spectra for aqueous, monodentate, and bidentate sulfate species. Adapted from Peak D, Ford RG, and Sparks DL (1999) An *in situ* ATR-FTR investigation of sulfate bonding mechanisms on goethite. *Journal of Colloid Interface Science* 218: 289–299.

covalently bound to sulfate (as bisulfate or sulfuric acid), but also when hydrogen bonds are formed with waters of hydration. To improve peak assignment in such cases, experiments are usually repeated in D_2O as a solvent. Substitution of D for H in surface complexes or mineral structures will shift the position of infrared peaks, while peaks that do not involve protons are unaffected. It is also possible to calculate theoretical infrared spectra using molecular modeling software. The absolute position of peaks may not be correct, but trends in how species are affected by the presence of protons or metals are still quite helpful in peak assignment.

FTIR as a Quantitative Tool

The infrared absorbance of a molecular vibration obeys the Beer–Lambert law. This means that FTIR spectroscopy, if properly calibrated, is a quantitative technique. The relationship is as follows:

$$A = \epsilon b C \qquad [1]$$

where A is absorbance, ϵ is the molar absorptivity coefficient, b is the thickness of the sample (or pathlength), and C is the concentration of the molecule in the sample. Note that ϵ is specific to a given vibration, and that A is typically reported in terms of peak height or integrated absorbance from a fitted peak. Changes in sample configuration for different types of FTIR experiments can make quantification difficult.

Instrumentation

FTIR actually refers to a class of spectrometers that are distinguished from older, double-beam infrared units by their use of a single beam and an interferometer and detector to collect a spectrum over the entire wavelength of interest simultaneously. This raw data, called an interferogram, is an oscillation over time. This interferogram is then converted to a function of frequency, as in a traditional infrared spectrum, by use of a mathematical procedure called Fourier transformation. FTIR units have largely replaced older double-beam units because they are have better spectral resolution, allow for much more rapid data collection, and have an improved signal-to-noise ratio.

Common Experimental Techniques

Several different FTIR experimental techniques are commonly employed by environmental researchers. These are transmission mode experiments with pressed disks, diffuse reflectance experiments with powder samples, and attenuated total reflectance studies with aqueous solutions and hydrated suspensions. The primary differences in the commonly used techniques are: (1) the amount of sample preparation that is required, (2) the spectral range that is accessible to the technique, and (3) the detection limits. Each technique may be particularly well suited to certain research questions, and more than one technique is sometimes required.

Transmission-mode Experiments

Transmission-mode experiments are the both the simplest and most common type of FTIR experiment. Samples are diluted in some medium and placed perpendicular to the infrared beam. A detector measures the amount of infrared beam that is transmitted through the sample. To obtain the spectrum of only the sample, the spectrum of the pure medium is first collected and used as a reference. A spectra of the sample in the medium is then collected and ratioed to the reference to obtain the spectrum of the sample. A typical experimental configuration for transmission-mode experiments is shown in Figure 3.

Solid samples are typically diluted to 1–2% wt in KBr salt and then made into a thin disk with a press. The final product is a thin, clear disk that is placed in a sample holder and analyzed. For transmission mode experiments to be successful, water must be removed from the system, so samples are heated to remove as much water as possible prior to dilution. Care must be taken with amorphous and unstable solids to avoid transformations that may occur at elevated temperatures. Organics can also often be studied in transmission mode by simply placing them in a thin liquid cell. The medium chosen as a solvent should have a relatively low infrared absorbance and not have infrared peaks that overlap with those of the organic of interest.

Transmission-mode experiments can provide quantitative information more easily than some other experiments because Beer's Law is directly applicable. The path length through the sample must be known, and molar absorptivity can be learned by collecting several spectra of samples with known concentrations to construct a calibration curve. Once the path length and molar absorptivity are known, concentration and absorbance can be determined.

This technique is quite useful for soil scientists who are studying mineral transformations or who need quantitatively to determine the mineral composition in unknown samples. However, the technique involves substantial sample modification that may change the structure of both solid phases and sorbed surface species. Particularly in mixed systems, peaks of minor components may be masked by other species of higher concentration.

Diffuse Reflectance Infrared Fourier Transform Experiments

The diffuse reflectance technique was developed partly to address some of the shortcomings of the pressed-disk method of sampling, and also to extend FTIR spectroscopy to additional types of samples. Many substances, particularly powders, scatter light in all directions in a diffuse manner. The diffuse reflectance technique isolates the portion of the infrared light that is diffusely scattered from that which is either transmitted or specularly reflected and collects it as a diffuse reflectance spectrum. A typical experimental configuration for transmission-mode experiments is shown in Figure 3.

DRIFT spectroscopy is conducted only on powder samples. Samples may be run neat (with no dilution),

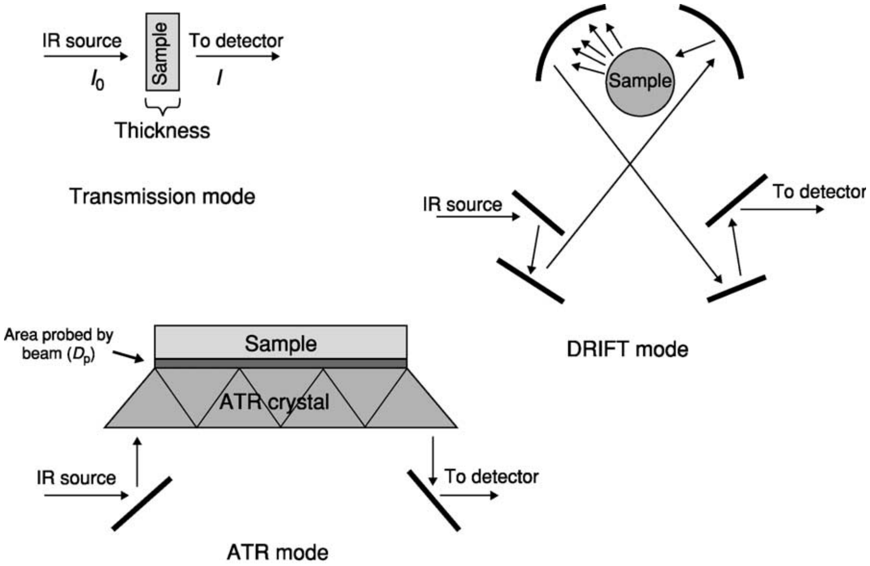


Figure 3 Common sampling methodologies for transmission-mode, diffuse reflectance (DRIFT), and attenuated total reflectance (ATR) Fourier transform infrared experiments. I_0, energy prior to the sample; I, energy after the sample; D_p, depth of penetration through the sample.

or diluted with a nonabsorbing matrix such as KBr. Samples are typically ground with a mortar and pestle to avoid large particles with flat surfaces that may distort the spectrum by exhibiting specular reflectance. Samples are not pressed into disks as with transmission experiments, but are simply packed into a sample holder. This makes the collection of DRIFT spectra somewhat simpler than transmission-mode experiments. It also avoids changes in sample structure that may occur when subjected to elevated pressure. However, quantitative DRIFT studies are more difficult than transmission-mode IR, because steps in sample preparation can substantially affect a diffuse reflectance spectrum. Specifically, the particle size, the sample homogeneity, and the degree of packing will influence the results.

Regardless of the care taken in sample preparation, DRIFT data differ from data obtained in a transmission-mode experiment. Peak intensities at higher wave number are typically compressed, and all peak shapes tend to be rather rounded compared with sharp peaks from pressed disks. A mathematical correction can be applied to the raw spectra to compensate for these anomalies and make DRIFT data directly comparable with transmission-mode references. This conversion is known as the Kubelka–Munk equation, and it can usually be performed by the software used to control the FTIR.

The primary advantage of DRIFT over transmission-mode studies is that sample preparation does not require pressing disks. This means samples can be prepared more quickly and with less effort. However, samples are still dried and diluted in salts, so sample preparation may still affect the chemical species in an environmental sample. The primary drawback of DRIFT is that it trades ease of sample preparation and ability to observe weaker bands by analysis of neat samples for additional complexity in quantitative analysis.

Attenuated Total Reflectance Experiments

The attenuated total reflectance (ATR) method is a recent advancement in FTIR that allows spectra to be collected in the presence of water. Water is an extremely strong infrared absorber, and traditional sampling methods require drying to remove interferences from H_2O. The ATR technique circumvents this limitation by directing the infrared beam through a long crystal. The beam is redirected through the crystal at an angle and bounces multiple times internally and then exits to the detector. The infrared beam produces an evanescent wave on the surface of the crystal, and hydrated samples in close contact with the crystal can be analyzed by comparing the reflectance of the sample with that of the bare crystal. A typical experimental configuration for ATR-mode experiments is shown in Figure 3.

Since the ATR technique presents a substantial deviation from a transmission-mode experiment, experimental complications that affect the adherence of ATR data to the Beer–Lambert law are to be expected. For example, the effective path length of the beam is required for quantitative research and is related to the number of reflections through the crystal and the depth of penetration into the sample. The number of reflections is related to the ATR crystal's physical properties such as angle of incidence of the crystal, and the crystal's physical length and thickness. The depth of penetration into the sample is also related to the crystal's physical dimensions, but is also a function of both the crystal and the sample's refractive index and the wavelength of light that is passed through the crystal. ATR crystals may have 30°, 45°, or 60° effective angles of incidence, and may be made of a variety of materials. The most common and widely used material for ATR crystals is ZnSe, but diamond, amorphous material transmitting IR (AMTIR), germanium, and silicon are also available. Different crystal materials may be preferred due to their refractive index, their useful spectral range, or their chemical sensitivity.

The fact that sample penetration depth changes with wavelength presents a serious problem in attempting quantitative experiments with the ATR method. For this reason, ATR data are properly reported as percentage reflectance (%R) rather than percentage transmittance (%T), as the intensity of light transmitted is not a constant throughout the spectrum with ATR, but instead is a function of wavelength. Fortunately, a mathematical correction can be made to ATR data that accounts for the wavelength dependence of penetration depth. This correction is typically available in the instrument's data-collection software.

With the ATR technique, it is possible to analyze a wide range of samples. Solids can be pressed into contact with the ATR crystal, while liquid samples and hydrated suspensions can be placed into a sample holder that has a trough with a volume of a few milliliters. This flexibility in sampling has made it possible to conduct adsorption studies of organic acids and oxyanions *in situ* with FTIR. In fact, the ability to monitor the mechanism of adsorption on mineral surfaces is the primary application of ATR-FTIR in environmental science.

Two experimental approaches are commonly utilized in adsorption studies: analysis of reacted suspensions of soil solids, and analysis of adsorption on to solids deposited on to the ATR crystal. Both experimental approaches have merits and drawbacks. It is faster and easier to conduct macroscopic adsorption experiments in the laboratory and then use ATR-FTIR spectroscopy to analyze representative samples than it is to conduct all experiments *in situ* with deposited solids. This also provides the opportunity to know the exact surface loading and reaction conditions of the sample prior to analysis. However, using wet pastes or slurries presents some experimental difficulties also. To obtain a spectrum of the adsorbed species, one must separate the solid from the solution via filtration, resuspend the solid in the supernatant at a constant solid-to-solution ratio for all samples, and then collect a spectrum of the supernatant, the unreacted solid, and the reacted solid. By subtracting both the supernatant and the unreacted solid spectra from the reacted spectrum, the result is the spectrum of the adsorbed species.

In contrast, the deposition method places a thin coating of a mineral on the ATR crystal that is constant throughout the experiment. A reference spectrum of the ATR crystal, the mineral deposit, background electrolyte, and water is collected initially. Then, reactant is added to the solution at a low (micromolar) concentration. The reactant is concentrated at the mineral–crystal interface so that adsorbed species are above detection limits. Since the solid concentration at the surface is constant and the solution-phase reactant is not detected, a high-quality spectrum of the absorbed ion can be obtained. Due to improved spectral quality, results from deposition studies are typically much easier to fit and interpret.

The primary advantage of the ATR method for environmental science research is that it allows environmental samples to be analyzed with normal amounts of water present and at ambient pressures. The results from ATR are therefore much more representative of chemistry in natural systems. However, ATR experiments are more difficult to compare with traditional transmission studies, difficult to conduct reliably in a quantitative manner, and the experimental procedure is more time-consuming than other methods.

Further Reading

Farmer VC (1974) *Infrared Spectra of Minerals*. London, UK: Mineralogical Society.

Johnston CT, Farmer W, and Aochi Y (1996) Fourier transform infrared and Raman spectroscopy. In: Sparks D (ed.) *Methods of Soil Analysis, part 3. Chemical Methods*, 3rd edn, pp. 269–321. Book Series no. 5. Madison, WI: Soil Science Society of America.

Nakamoto K (1986) *Infrared and Raman Spectra of Inorganic and Coordination Compounds*. New York: John Wiley.

Noggle J (1998) *Physical Chemistry*. New York. NY: Harper-Collins.

Shriver DF, Atkins PA, and Langford CH (1994) *Inorganic Chemistry*. New York, NY: WH Freeman.

Suarez DL, Goldberg S, and Su C (1998) Evaluation of oxyanion adsorption mechanisms on oxides using FTIR spectroscopy and electrophoretic mobility. *ACS Symposium Series. American Chemical Society* 715: 136–178.

Wilson EB, Decius JC, and Cross PC (1980) *Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra*. New York, NY: Dover Publications.

FRACTAL ANALYSIS

Y Pachepsky, USDA Agricultural Research Service, Beltsville, MD, USA
J W Crawford, SIMBIOS, University of Abertay, Dundee, Scotland, UK

Published by Elsevier Ltd.

Introduction

Geometric irregularity is an intrinsic property of soils. Pores, particles, surfaces, roots, and living organisms in soils have a wide range of sizes and shapes. Soil surface is rough, and soil map contours have complex shapes. Geometric irregularity in soils is easy to perceive and observe, but quantifying it has long presented a daunting challenge. Such quantifying becomes imperative because the geometric irregularity is both a cause and a reflection of soil spatial variability that strongly affects soil functioning in ecosystems and soil management.

Geometric quantities in soils are measured with different measurement scales, i.e., lengths, areas, or volumes within which curves or surfaces are assumed to have a geometric shape, composed of straight lines, circular arcs, parts of planes, or sphere surfaces. Measurement scale is often called ‘support,’ ‘resolution,’ or simply ‘scale.’ Because actual curves and surfaces are irregular within the measurement scale, total length, area, or volume does not remain

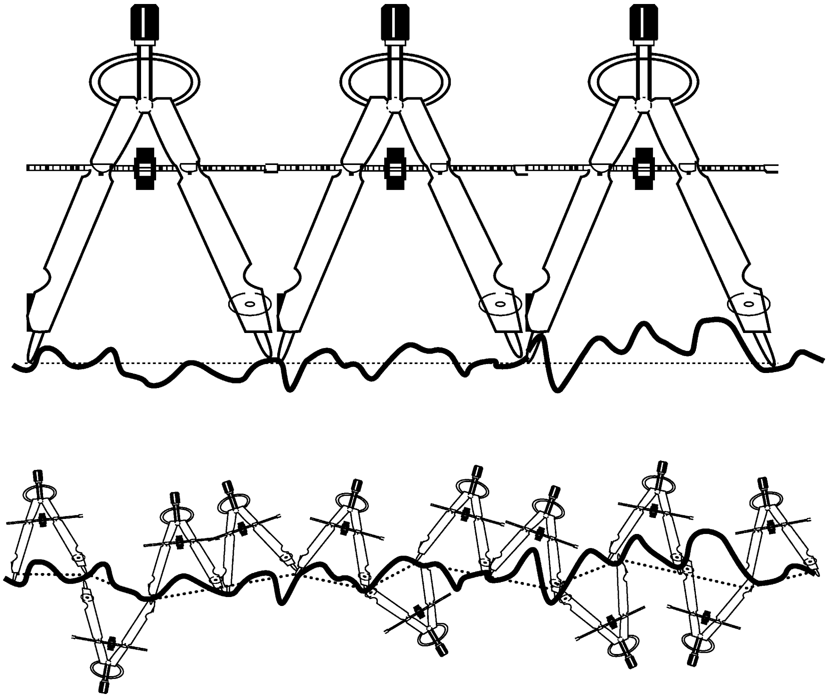


Figure 1 Measurements with traditional geometric tools ignore surface irregularity within the observation scale. Measured length of a line (dotted) is the sum of distances between the compass dividers. The larger divider step gives the total length of 105 length units, whereas the smaller divider step results in 120 length units.

constant as the measurement scale changes (Figure 1). A decrease in measurement scale usually causes an increase in total length, area, or volume. Thus soil geometric properties appear to be scale-dependent.

One class of geometric objects that have dependence of total length, area, or volume on the measurement scale consists of objects that reveal similar geometric shapes when observed at different measurement scales. Such objects were termed 'fractals' by B.B. Mandelbrot, who suggested that fractals rather than regular geometric shapes such as segments, arcs, circles, and spheres are more appropriate to approximate irregular natural shapes that have hierarchies of ever-finer detail. This started the expansion of applications of fractal geometry, which had became very popular during recent decades, because of its promise to relate features of natural objects observed at different scales.

Fractal Objects

Geometric fractal objects reveal similar features at different scales because they are constructed iteratively, by repetition of the same form- or shape-changing operation at different measurement scales. If this operation does not change from one scale to another, then the result is an ideal geometric fractal. An example of such construction is given in Figure 2 for the fractal object 'Menger sponge.' The initial shape, or initiator, is a cube. The shape-changing operation, or generator, divides the cube into 27 smaller, equal cubes and extracts seven cubes, one from each facet and one from the center. Figure 2 shows the first two stages of the iterative application of this operation. A Menger sponge is self-similar because it can be broken down into arbitrary small pieces, each of which is a small replica of the entire structure. If the Menger sponge of a certain size is reduced three times, then 20 copies of the reduced sponge can be pasted together to give back the original Menger sponge; a nine-times reduction requires 400 copies; a 27-times reduction requires 8000 copies, etc. In all self-similar objects, there is a relationship between the reduction factor R and the number of pieces N into which the object is divided. Those relationships follow a power law:

$$N = R^{-D_{ss}} \qquad [1]$$

where the exponent D_{ss} is called the self-similarity, or similarity, dimension. The value of D_{ss} can be found by equating logarithms of both parts in Eqn [1], i.e., $D_{ss} = -\log(N)/\log(R)$. For the Menger sponge, the reduction factor is one-third $D_{ss} = -\log(20)/\log(1/3) = -\log(400)/\log(1/9) = -\log(8000)/\log(1/27) = 2.727$. The reduction factor refers to a single iteration; the total reduction after i iterations is $(1/3)^i$.

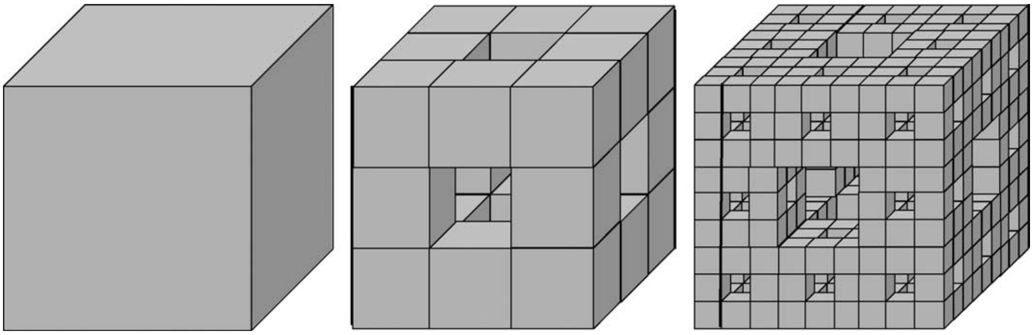


Figure 2 The first two iterations of the Menger sponge construction.

The self-similarity dimension receives a geometric meaning when a fractal object is used as the model of a real system, and geometric measures of this object are equated to geometric measures of the system. In this case, the measurement scale is the reduced length, area, or volume of the initiator. For example, let empty parts of the Menger sponge be used to simulate pores in a porous medium at different measurement scales. Let V_0 be the volume of the initiator. The measurement scale is V_0R^3, the number of similar pieces at this scale is N, and their total volume is:

$$V = NV_0R^3 = V_0R^{3-D_{ss}} \qquad [2]$$

As the measurement scale becomes finer and R decreases, the remaining solid volume decreases, approaching zero, and the pore volume approaches V_0. Because the fractal object represents remaining mass in this case, the dimension D_{ss} is interpreted as the mass fractal dimension. Depending on the scale-dependent measure of interest represented by the specific fractal object or on a physical process simulated by the generator, D_{ss} is interpreted as being the surface fractal dimension, fragmentation fractal dimension, etc. Eqns [1], [2], and similar relationships hold over a range of scales, i.e., they are scale-invariant within this range. They are interchangeably referred to as 'scaling relationships,' 'fractal scaling laws,' or simply 'scaling.'

A common way to reveal a fractal scaling is to plot the measurement scale against the property of interest in double-logarithmic, or log-log, scale. The presence of a linear relationship indicates that a fractal model may be appropriate to simulate the scale dependence in the property. The fractal dimension can then be found from the slope of the regression line. For example, Eqn [2] in log-log scale can be transformed into:

$$\log V = \log V_0 + (3 - D_{ss})\log R \qquad [3]$$

and the fractal dimension D_{ss} equals 3 minus the slope of the regression line.

The iterative construction of ideal fractals continues ad infinitum, whereas natural objects have a minimum and maximum observable scales, or lower and upper cutoff scales beyond which Eqn [1] breaks down. The value of the measure for the initiator, for example the V_0 value in Eqn [2], is not known if the fractal scaling is only valid within a specific scale range. Thus, four values – fractal dimension, upper and lower scale cutoff, and the initiator measure – are parameters of a fractal model in soil studies. Usually, the range of scales between cutoffs does not exceed two orders of magnitude if scales are expressed in units of lengths. The larger the fractal dimension the more rugged is the surface, the more irregular is the line, and the more complex is the pore space.

Cutoffs typically coincide with changes in processes causing or reflecting scaling; for example, pore-space fractal scaling inferred from water retention data in the capillary range is limited by the water content at wilting point and the water content at which the macropores become emptied. The meaning of the initiator measure is not currently well understood, although concepts of topothesy, crossover length, and lacunarity have been explored to interpret this fractal parameter.

Fractal dimensions of geometric fractals have limits reflecting space-filling properties. Straight line, plane, and Euclidean space have dimensions one, two, and three, respectively. A fractal line has its fractal dimension between one and two, i.e., it does not fill the whole plane although the larger the fractal dimension is the more rugged is the line and the closer it is to occupy a part of the plane. Similarly, a fractal surface has its fractal dimension between two and three. Fractal dimensions of surface and volume fractals theoretically cannot exceed three.

Ideal geometric fractals as in **Figure 2** cannot be found in real porous media in the same way as cannot be found ideal geometric lines, spheres, cylinders, etc. Applying random changes in the generator procedure results in constructing stochastic fractals that look much more like fractals in nature. Because random changes are applied, various realizations of stochastic fractals with the same initiator and generator differ. Stochastic fractals retain the important feature of dependence of geometric properties on a measurement scale, only this feature is preserved in a statistical sense. This means that Eqn [1] will not hold exactly for any realization of stochastic fractals at any measurement scale. Nevertheless, a plot of the measurement scale versus the property of interest in log-log scale is close to linear for each realization of a stochastic fractal, and statistical hypotheses of equality between average slope values and average intercept values for any two realizations cannot be rejected. Realizations of stochastic fractals are statistically similar.

Data on surface roughness, topography, and distributions of properties along spatial transects do not follow the self-similarity law (Eqn [1]), but demonstrate another, anisotropic type of scale-dependence, 'self-affinity.' Self-similar fractals are isotropic, which means that, in two-dimensional xy space, $f(x,y)$ is statistically similar to $f(Rx,Ry)$ where R is a reduction factor. Self-affine fractals are anisotropic, which means that in the two-dimensional xy space, $f(x,y)$ is statistically similar to $f(Rx,R^{\mathrm{H}}y)$ where H is called the Hurst exponent. That is, changing the horizontal measurement scale R times corresponds to changing the vertical measurement scale R^{H} times. High values of the Hurst exponent indicate some memory or autocorrelation in the data. Low values suggest an anticorrelation or self-correcting response. Self-affine surfaces are observed in three dimensions, so that the vertical scale has to be decreased R^{H} times and scales in horizontal directions have to be decreased R times to obtain a statistically similar surface. The fractal dimension of this surface is $D_{\mathrm{s}} = 3 - H$. A line formed as a cross-section of any fractal self-affine surface and a vertical plane also displays self-affine properties. This line will have a fractal dimension D_{L}, related to the Hurst exponent H as:

$$D_{\mathrm{L}} = 2 - H \qquad [4]$$

The value of D_{L} is easy to interpret with elevation data. Any two consecutive height variations are likely to have opposite signs when the fractal dimension D_{L} is larger than 1.5 at any particular scale. Any two consecutive height variations are likely to have the same sign when the fractal dimension D_{L} is less than 1.5 (Figure 3).

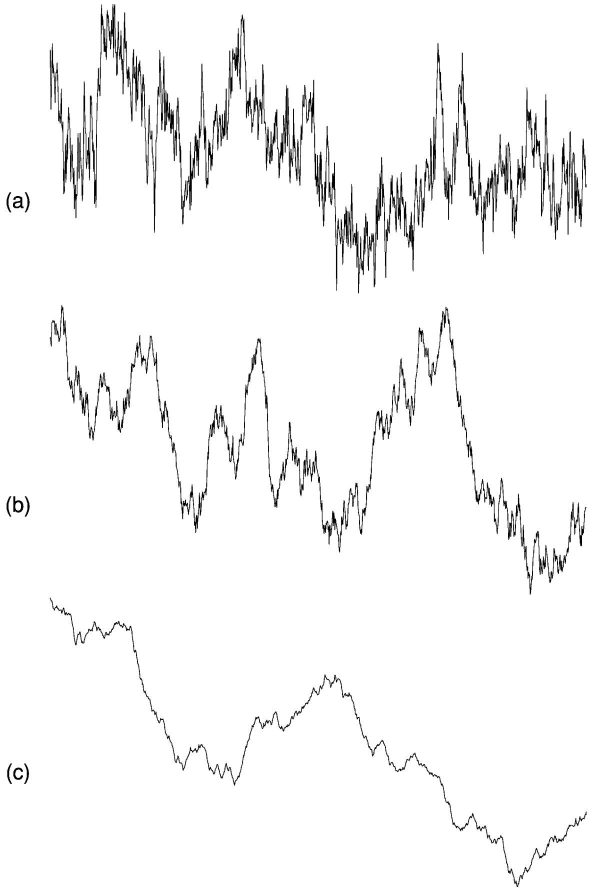


Figure 3 Self-affine fractal lines: (a) fractal dimension $D_{\mathrm{L}} = 1.8$; (b) $D_{\mathrm{L}} = 1.5$; (c) $D_{\mathrm{L}} = 1.2$.

Multifractals present yet another example of self-similar geometric objects that have been successfully applied to simulate properties of soils and other natural systems. Unlike the Menger sponge, in which the similarity of structure persists across iterative steps, multifractals are generated by cascade processes that create more heterogeneity as the resolution decreases. An example of a cascade process is shown in Figure 4. A unit-length segment is the initiator. The reduction factor is 2, and random split of each segment from the previous step into two new segments occurs as the generation progresses. The random-split variable ranges from 0 to 1, with a mean value of 0.5. The distributions of lengths are monofractal if the distribution of the random-split variable does not change from one scale reduction to another. However, if the variance of the random-split variable decreases as the generation continues, the distributions become multifractal. Distinguishing between monofractal and multifractal distributions can be based on considering statistical moments about zero:

$$E[W^{\mathrm{q}}] \cong \frac{\sum_{i=1}^{m} W_{\mathrm{i}}^{\mathrm{q}}}{m} \qquad [5]$$

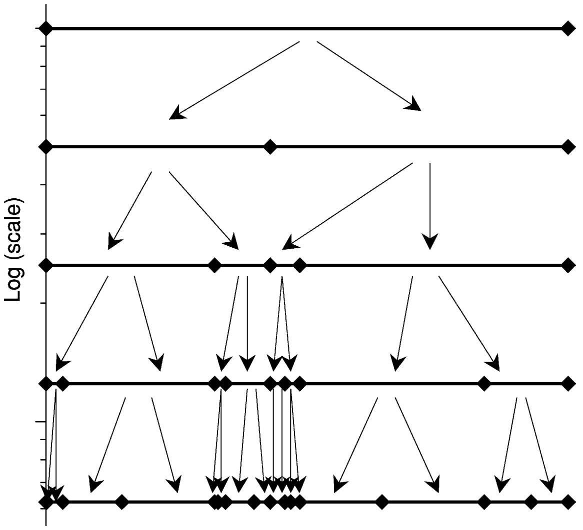


Figure 4 Generating multifractal-size distribution with a cascade process.

where q is the order of the statistical moment, W is the scale-dependent measure, and m is the total number of measurements. Multiscaling requires that:

$$E\left[W_{\mathrm{R}}^{\mathrm{q}}\right] = R^{\mathrm{K(q)}} E[(W_0)^{\mathrm{q}}] \qquad [6]$$

where K is the scaling exponent, W_0 is the measure at the measurement scale zero, and W_{R} is the same measure at the scale changed R times. If $K(q)$ is proportional to the moment order q, then the scaling is called 'simple scaling.' If $K(q)$ is a nonlinear function of q, scaling is interpreted as multiscaling or multifractal.

Various other types of fractal objects can be generated that are suitable to simulate scaling in soil properties. Scaling exponents and fractal dimensions have different meanings for those objects, because the construction algorithms are different.

Physical Models of Emerging Fractal Scaling

Many physical processes have been shown to generate fractal scaling. Some of them are relevant to soil development and functioning, and the correspondent fractal models are being applied to soils. One example is the scale-invariant fragmentation that leads to dependence between the number of fragments and their size:

$$N_{\mathrm{r}} = kr^{-D_{\mathrm{F}}} \qquad [7]$$

where r is the fragment radius, N_{r} is the number of fragments having radii less than r, k is a constant, and D_{F} is the fragmentation fractal dimension. The incomplete fragmentation process has the fractal dimension that can be related to density-size scaling of soil aggregates and appears to be the mass fractal dimension.

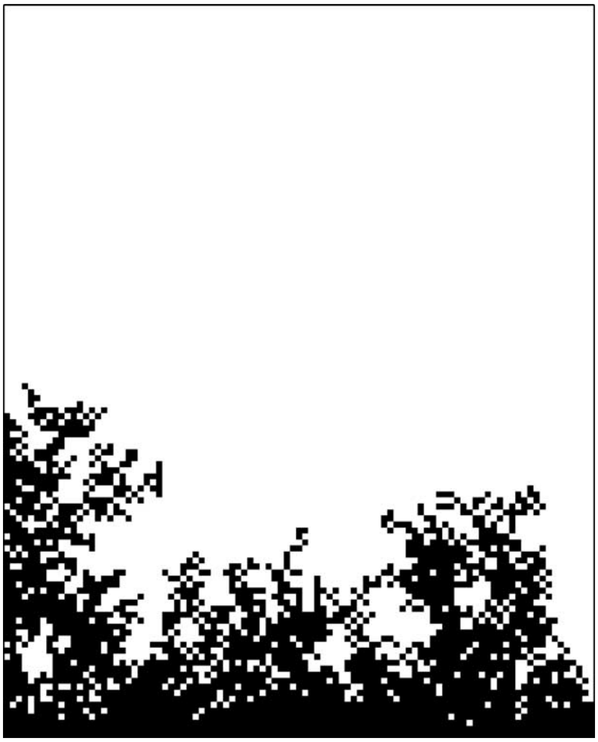


Figure 5 Diffusion-limited deposition of particles on a surface.

Agglomeration on a substrate can also generate fractal structures. Particles may arrive at the surface by random diffusion or with ballistic energy. They may either just stick together as they touch each other or there is a chemical reaction involved which may change the local concentration gradient and sticking probability. Allowing particles to walk randomly until they strike the surface or another particle, which is already motionless, builds the fractal coating. An example of diffusion-limited deposition on a substrate is shown in Figure 5. Clusters are formed, and the fractal dimension of such clustering can be measured by log-log plotting the total number of particles as a function of distance from the surface. The fragmentation dimension appears to be close to the surface fractal dimension computed for pores if particles have a fractal-size distribution following the fragmentation law (Eqn [7]) and their sedimentation is simulated as a ballistic process of agglomeration on a substrate (Figure 6). The number of particles in a cross section of fractal agglomerate on a surface can be interpreted as a solute concentration, and the agglomerate can be viewed as a model for the distribution of solute particles in a porous medium (Figure 7).

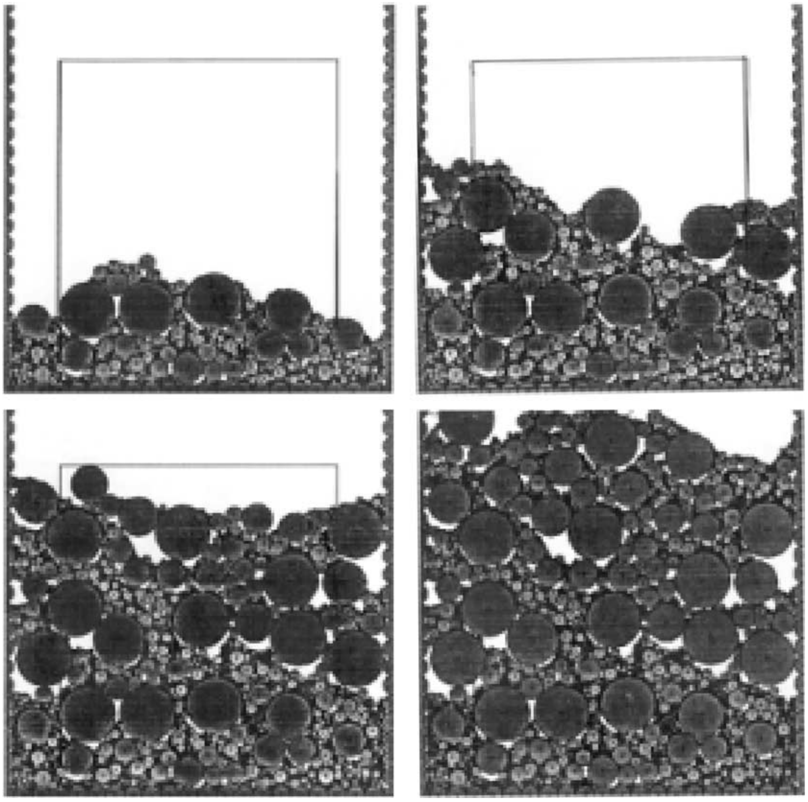


Figure 6 Consecutive stages of particle sedimentation simulated with a ballistic deposition algorithm; particle-size distribution was derived from the fractal fragmentation model. Reproduced with permission from Ghilardi P, Kai AK, and Menduni G (1993) Self-similar heterogeneity in granular porous media at the representative elementary volume scale. *Water Resources Research* 29: 1205–1214.

Random motions of large numbers of particles can cause fractal scaling in the particle ensemble characteristics. In particular, the fractal Brownian motion can be an underlying generator of data sequences that demonstrate self-affine scaling with the Hurst exponent $H \neq 0.5$ and $D_L \neq 1.5$ in Eqn [4]. Values of $H < 0.5$ are often related to the presence of correlated structural units in soil structure at different scales.

Measuring Fractal Parameters of Soils

Both direct and indirect measurements of geometric quantities are in use in soil studies. Indirect, or proxy, geometric measurements are very common. Nongeometric values are converted into geometric values using a physical model thought to be approximately valid in soils. For example, capillary pressures are converted into radii using capillary models, masses of adsorbed molecules are converted into area values using a molecular monolayer model, sedimentation time is converted into particle radii using the Stokes law, etc. With both direct and proxy measurements, methods of determining fractal parameters differ mostly in the way of changing the measurement scale.

Direct Geometric Measurements

The divider method is the first method applied to estimate the fractal dimension of self-similar lines in nature. The compass dividers are set to a specific stride length (measurement scale) as shown in Figure 1, and then the line is walked along. The fractal dimension is one minus the slope of the log(length)-log(scale) plots. Figure 8 shows boundaries of the fingers appearing at the wetting front in layered model soil systems that have been measured with various stride lengths, and the values of fractal dimension were successfully related to the distance of the finger penetration.

The box-counting method is based on covering the studied area with grids of different mesh sizes Δ and, for each mesh size, counting the number of grid squares N_B that include the studied line. The dependence between N_B and Δ follows a power-law relationship:

$$N_B \propto \Delta^{-D_{L,B}} \qquad [8]$$

if the line is fractal; and $D_{L,B}$ is the Kolmogorov dimension which approximates the fractal dimension of the line. The box-counting method is applicable to complex outlines and to branching lines (Figure 9).

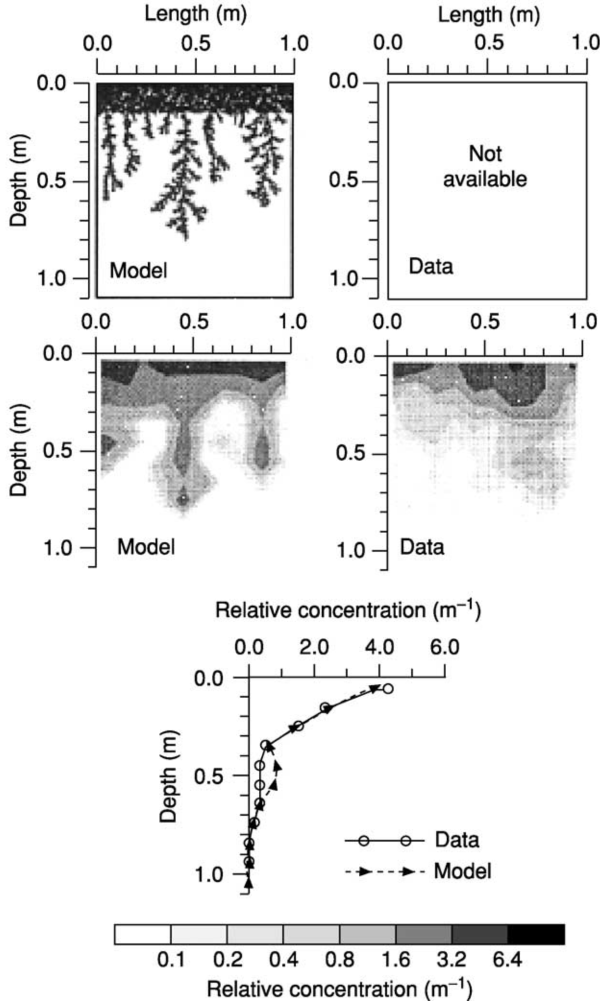


Figure 7 Comparison between the diffusion-limited aggregation model predictions and experimental data on bromide transport in layered soil. No comparison can be made at a microscopic scale of 0.2 cm representing the grid at which particles have traveled in simulations. At the field observation scale of approximately 10 cm, random distributions of solute concentrations in simulations and in experiment are difficult to compare. Mean concentrations averaged horizontally across the 1.4 × 1.4-m plot compare satisfactorily. Reproduced with permission from Flury M and Flühler H (1995) Modeling solute leaching in soils by diffusion-limited aggregation: basic concepts and application to conservative solutes. *Water Resources Research* 31: 2443–2452.

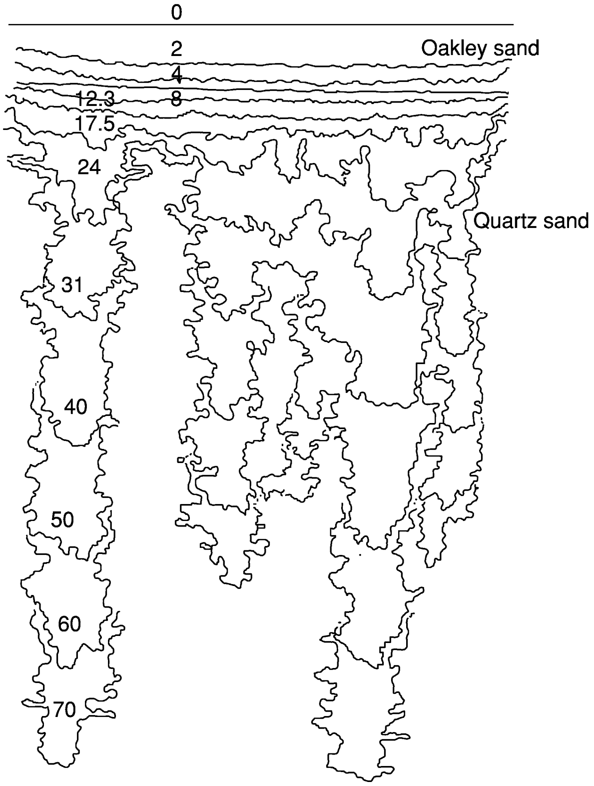


Figure 8 Infiltration and movement of water in quartz sand beneath Oakley sand. Each contour outlines the wetted area. Numbers at the contours show the time from the beginning of infiltration (minutes). Measurement of the contour lengths has been made with the divider. The divider stride length was varied from 1 to 18 cm for each of the contours, and the line fractal dimensions were computed from relationships between the measured contour length and the divider stride used as the measurement scale. Values of fractal dimensions were in the ranges from 1.007 to 1.09 and from 1.31 to 1.42 before and after the front reached the boundary between layers, respectively. Reproduced with permission from Chang WL, Biggar JW, and Nielsen DR (1994) Fractal description of wetting front instability in layered soils. *Water Resources Research* 30: 125–132.

Fractal dimensions can also be computed for lines defined as boundaries in images originating, for example, from soil thin sections or from remote sensing data. There is more uncertainty in defining fractal parameters with data on lines from imagery, because lines are defined by separation of two phases. Separation of two phases is based on a subjectively selected threshold between two ranges of brightness, and a noise which is always present in imagery can result in a rougher apparent boundary. Box-counting is applicable to such data. Progressive coarsening of images is yet another way to observe the line at different scales and to compute the Kolmogorov dimension.

Erosion- and dilation-based techniques are also used with lines from images. Dilation adds any background pixel that touches the line, while erosion removes any pixel that touches the background. Performing each of these operations M times and counting pixels that are affected as a function of M produces a plot whose slope in log-log coordinates gives the Minkovsky dimension that approximates the fractal dimension D_L.

Self-affine lines require different techniques to estimate the Hurst exponent and the fractal dimension. The Korcak method consists in making a cumulative plot of the number of intercepts as a function of

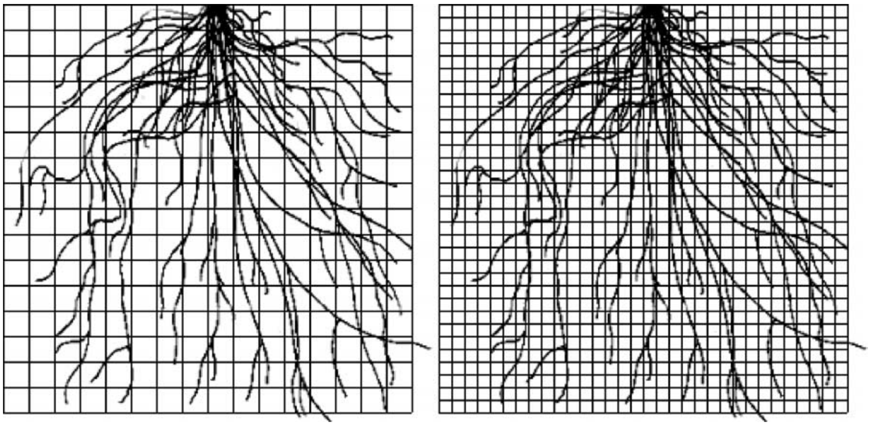


Figure 9 Box-counting to estimate fractal dimension of the root system for a maize plant grown in boxes 1.2 m long, and 0.05 m wide. Roots cross 198 grid cells in the 16 × 16 grid and 554 cells in the 32 × 32 grid. From those two measurements, the Kolmogorov dimension is $D_{L,B} = -\log[N_B(16)/N_B(32)]/\log[(1.2/16)/(1.2/32)] = -\log(198/554)/\log(2) = 1.484$. Reproduced with permission from Eghball B, Settimi JR, Maranville JW, and Parkhurst AM (1993) Fractal analysis for morphological description of corn roots under nitrogen stress. *Agronomy Journal* 85: 287–289.

lengths for any arbitrarily selected horizontal line. A log-log plot of the number of distances that exceed a value x as a function of x has a slope that is exactly 1 less than the fractal dimension of the profile. The semivariogram method is based on quantifying the spatial structure in a self-affine line using the dependence between variability in measurements and the distance between the measurement points. The semivariogram γ is computed as:

$$\gamma(h) = \frac{1}{2m_h}\sum[Z(x) - Z(x+h)]^2 \qquad [9]$$

where Z is the observed point on the self-affine line, e.g., elevation, h is the distance between observation points, interpreted as a measurement scale in this case, and m_h is the total number of pairs of points separated by the distance h. For a self-affine line:

$$\gamma \propto h^{2H} \qquad [10]$$

and the log-log plot of h versus γ provides the value of H and D_L from Eqn [4], as shown in Figure 10.

The power-spectrum method is based on Fourier analysis. Squared coefficients of the Fourier expansion, or magnitudes, are plotted against frequencies in log-log scale. The slope β of the linear regression line is related to the fractal dimension of the line as $D_L = 2 + \beta/2$. This method can also be applied to measure the fractal dimension of surfaces D_s, as $D_s = 3 + \beta/2$ (Figure 11).

Roughness-based methods are also applied to estimate fractal parameters of lines and surfaces. In the absence of a spatial trend, variances of the property in question are computed for different measurement scales. This is best done by placing many windows of the same width w (or area) on the data, either systematically or randomly, and calculating the mean value of the standard deviation for multiple windows. The procedure is repeated with windows

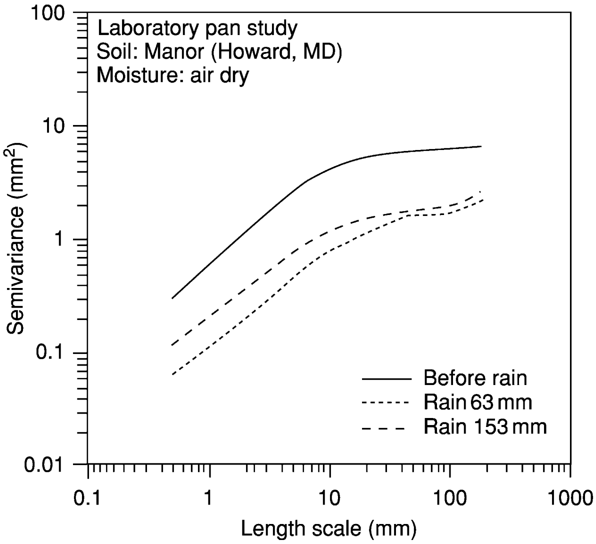


Figure 10 Changes in the semivariogram of soil surface elevation and in surface fractal dimensions caused by simulated rainfall. The tilted laboratory pan was filled with surface soil (coarse loamy, micaceous, mesic Typic Dystrochrept), and the surface topography of the central pan area of 0.54 × 0.6 m^2 was digitized with laser scanner before and after rain events. Surface fractal dimensions are 2.51, 2.83, and 2.68 before rain, after 63 mm of rain, and after 155 mm of rain, respectively. The lower and the upper cutoff values of the fractal scaling in this case are 0.6 mm and 10 mm, respectively. Reproduced with permission from Huang C and Bradford JM (1992) Applications of a laser scanner to quantify soil microtopography. *Soil Science Society of America Journal* 56: 14–21.

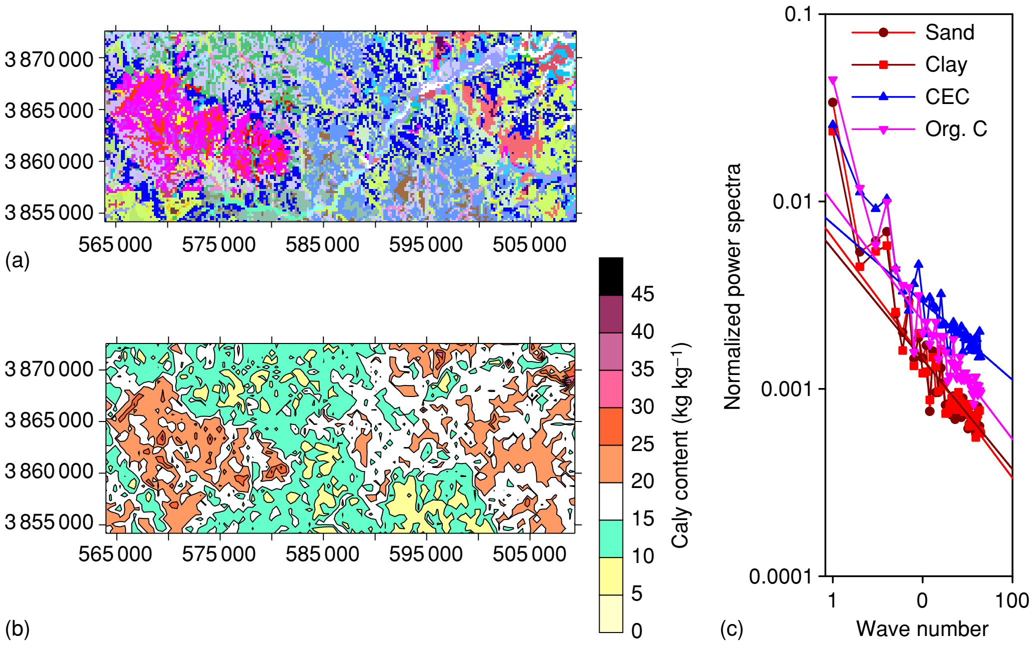


Figure 11 Application of the power-spectrum method to estimate surface fractal dimensions of soil basic properties in an area within the Little Washita watershed, OK, USA: (a) 44 individual soils defined from the MIADS data set for 200 m × 200 m pixels within the area; total number of pixels was 93 × 225; (b) map of clay content defined for each soil as the average value over the range given by soil survey; values of sand and organic carbon contents, and cation exchange capacity were defined for each pixel analogously; (c) power spectra developed for soil basic properties using their pixel values; the upper cutoff of fractal scaling is approximately 7 km, the lower cutoff is the pixel size of 200 m for all shown properties; estimated surface fractal dimensions are 2.68, 2.71, 2.79, and 2.68 for clay content, sand content, cation exchange capacity (CEC), and organic carbon (Org. C) content, respectively.

of different sizes, and the standard deviations are plotted against the window size in the log-log scale. The slope of the regression line added to 2 or 3 gives an estimate of the fractal dimension for the line or surface, respectively. One implementation of roughness-based methods consists in removing the linear trend in data within the windows and computing the root-mean-square (RMS) of residuals, or 'roughness,' as:

$$\mathrm{RMS} = \frac{1}{n_\mathrm{w}} \sum_{i=1}^{n_\mathrm{w}} \sqrt{\frac{1}{m_\mathrm{i} - 2} \sum_{j=1}^{m_\mathrm{i}} \epsilon_\mathrm{ij}} \qquad [11]$$

where n_w is the total number of windows of length w in the data, m_i is the number of points in the ith window, and ϵ_ij is the residual of the linear regression in the jth data point in the ith window. The window size serves as a measurement scale.

The area-perimeter, or slit-island, method can be used to find fractal parameters of complex outlines that can be seen in soil thin sections. This method does not require perimeter measurements with different step length but, instead, requires measurements of area and perimeters of the outlines. The fractal dimension of the irregular outlines, D_L, representing either particles or pores, is included in the relationship between area A and perimeter P:

$$P \propto A^{D_\mathrm{L}/2} \qquad [12]$$

Results of applying this relationship to thin-section images of soils under different long-term management are shown in Figure 12.

The number of visible macropores N_M that have radii larger than r often scales as $N_\mathrm{M} \propto r^{-D_\mathrm{M}}$, and fractal parameters of this scaling can be found from a log-log plot.

Proxy Measurements

Data on adsorption of molecules of different sizes reveal probing irregularity of surfaces at different scales, as shown in Figure 13. Assuming that molecules form a monolayer, one can use the molecule radius r as a measurement scale and relate the number of molecules N_m needed for monomolecular coverage of a fractal surface to the molecule radius as:

$$N_\mathrm{m} \propto r^{-D_\mathrm{s}} \qquad [13]$$

where D_s is the surface fractal dimension. The area of the monomolecular coverage of spherical particles is $A_\mathrm{m} = 4\pi r^2 N_\mathrm{m}$ and it changes with the radius of the

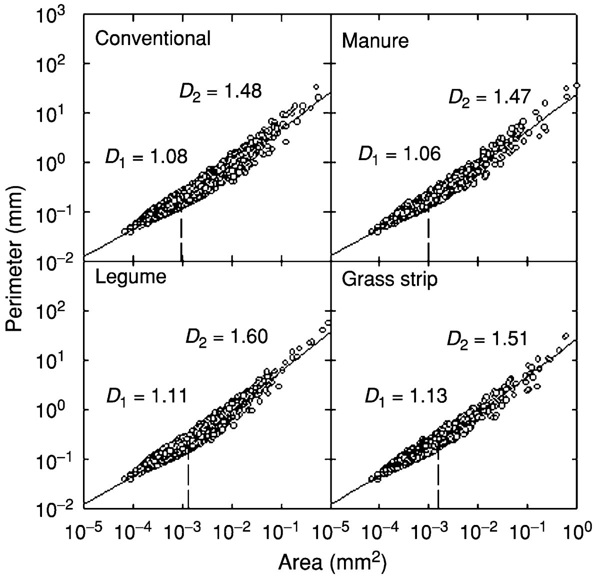


Figure 12 Fractal dimensions of pore outlines from thin section images for the Comly silty loam soil in plots where nitrogen was supplied from conventional fertilizer sources, manure, legumes, and in uncultivated plots under grasses where no nitrogen has been supplied. All plots were included in the experiment for 11 years. Small pores with areas less than 10^{-9} m^2 had fractal dimension D_1 between 1.06 and 1.12; pores with areas exceeding 10^{-9} m^2 had fractal dimension D_2 between 1.42 and 1.51. The value of D_2 is larger in samples from the legume plots than in samples from other plots. The cutoff between two linearity ranges corresponded to the boundary between vuggy and rounded pores. Reproduced from Pachepsky YA, Yakovchenko V, Rabenhorst MC, Pooley C, and Sikora LJ (1996) Fractal parameters of pore surfaces as derived from micromorphological data: effect of long-term management practices. *Geoderma* 74: 305–325.

molecule as $A_m \propto r^{2-D_s}$. The value of $D_s > 2$ and an increase in r is followed by a decrease in the area which is faster than r^2 because of the appearance of large caverns, as shown in Figure 13. The value of N_m can be estimated from the classic Brunauer–Emmett–Teller (BET) equation or from its modifications for fractal surfaces.

Isotherms of adsorption on fractal surfaces can be used to estimate fractal parameters when adsorption data for a single adsorbate are available in the range of relative pressures close to saturation. A fractal analog of the Frenkel–Halsey–Hill isotherm equation is:

$$n \propto \left(-\ln\frac{p}{p_0}\right)^{D_s-3} \qquad [14]$$

Here n is the adsorption and p/p_0 is the relative pressure. The measurement scale is the mean radius of the liquid–vapor interface, r_i, which is calculated from the Kelvin equation:

$$r_i = \frac{2\sigma V_m}{RT(-\ln\frac{p}{p_0})} \qquad [15]$$

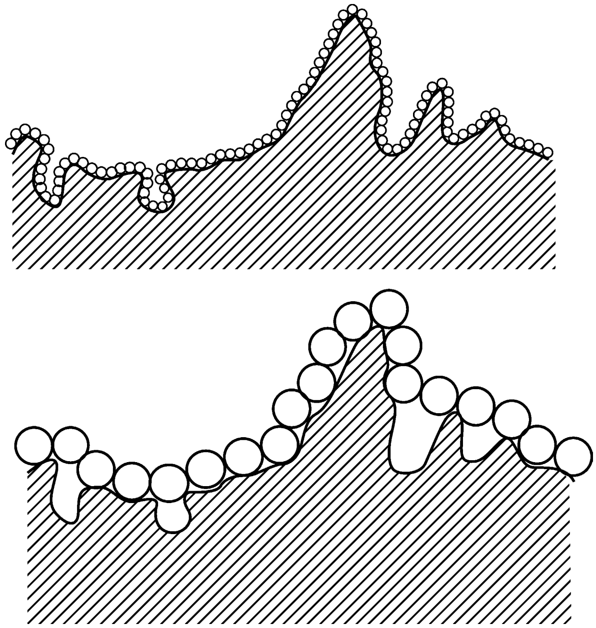


Figure 13 The molecular accessibility of a rough surface depends on the molecule size: larger molecules contact a surface that is much smaller than that contacted by smaller molecules.

where R is the gas constant, T is the temperature (Kelvin), σ is the surface tension of the liquid adsorbate, and V_m is the molar volume of the liquid adsorbate. Eqn [14] was derived by assuming the sequential filling of pores from small to large, that is, capillary condensation-type local adsorption isotherm. Eqn [14] is used with data obtained for $p/p_0 > 0.7$. A different model for calculating the fractal dimension D_s from the adsorption isotherm data reflects the idea that a surface is smoothed by the adsorbate, and the surface of the adsorbent covered with an adsorbate is smaller than the surface of the uncovered adsorbent. The area A_i of the 'condensed adsorbate–vapor' equilibrium interface is obtained as:

$$A_i = \frac{RT}{\sigma}\int_n^{n_{max}} (-\ln\frac{p}{p_0})\mathrm{d}n \qquad [16]$$

Here n_{max} is the maximum adsorption, n is the adsorption corresponding to the value of A_i. The fractal scaling law relates the area to the mean radius of the interface (Eqn [15]), and the resulting equation is:

$$A_i \propto \left(-\ln\frac{p}{p_0}\right)^{D_s-2} \qquad [17]$$

Both Eqn [14] and Eqn [17] are represented by straight lines in log-log coordinates, and slopes give values of the surface fractal dimension as shown in Figure 14. A caution has to be exercised in interpreting

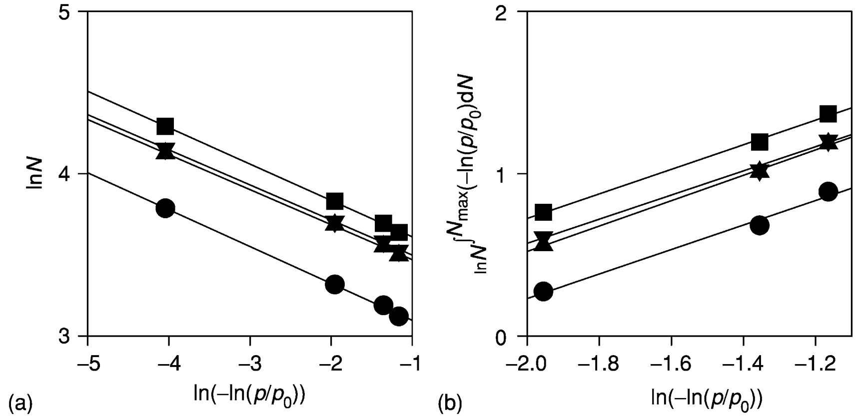


Figure 14 Fractal scaling in water vapor adsorption measured in samples of Udic Argiboroll, Moscow region, from four plots where a long-term experiment on grassing arable land had been carried out for 12 years: (a) Eqn [14], (b) Eqn [17] applied. Circles, no harvesting, no fertilization; squares, no harvesting, 60 kg NPK annually; triangles, harvested, no fertilization; inverted triangles, harvested, 60 kg NPK annually. Values of the surface fractal dimension D_S are in the range 2.75–2.85. Removal of carbohydrates has resulted in an increase in D_S. Differences in management practices did not affect values of D_S in the scale range studied, whereas the monolayer capacity was affected. Reproduced from Hajnos M, Korsunskaia LP, and Pachepsky YA (2000) Fractal dimensions of soil pore surfaces in a managed grassland. *Soil and Tillage Research* 55: 63–70.

values of fractal dimensions obtained with adsorption because of specific interactions of the adsorbate molecules with soil-pore surfaces.

The mercury porosimetry and water desorption are both used to estimate mass or surface fractal dimension of soil from the scaling relationship:

$$\frac{dV}{dr} \propto r^{2-D} \qquad [18]$$

Here V is the pore volume not yet filled by the intruding mercury or occupied by water, r is the mean pore radius serving as the measurement scale and estimated from the pressure values with the Laplace equation for capillaries. Selection of the geometric model of fractal medium determines whether D is an estimate of the surface or the mass fractal dimension.

Light, UV, and X-ray scattering have been applied to study fractal dimensions of organic substances in soils. Measurements with the single wavelength take advantage of the scaling dependence of the intensity of the scattered radiation I on scattering vector Q:

$$I(Q) \propto Q^{-s} \qquad [19]$$

where the scaling exponent s is equal to the mass fractal dimension D_m if the object is the mass fractal, and $s=6-D_s$ if the object of the study is the surface fractal. Measurements with multiple wavelengths relate the turbidity τ of a suspension to the wavelength λ as the measurement scale:

$$\tau \propto (\lambda)^{S-4+\delta} \qquad [20]$$

where S is the scaling exponent, δ is the value specific to the structure of organic molecules, $\delta=-0.2$ for humic substances. Values of S less (or greater) than 3 imply the presence of the mass (surface) fractal. Log-log plotting wavelength versus turbidity results in straight lines with slopes related to fractal dimensions, as shown in Figure 15.

Measurements of mass are used to estimate the number of particles in fragmentation scaling (Eqn [7]). The number of particles in the radius range r_i to r_{i+1} is found as $M/C\rho\bar{r}^3$, where M is the total measured mass of particles, ρ is the particle density, C is the constant reflecting the deviation of particle shapes from sphere, and $\bar{r}$ is the characteristic radius of particles in the range r_i–r_{i+1}:

$$\bar{r} = \begin{cases} \dfrac{D_F}{3-D_F}\dfrac{r_{i-1}^{3-D_F} - r_i^{3-D_F}}{r_i^{D_F} - r_{i-1}^{D_F}}, D_F \neq 3 \\ \dfrac{3\ln(r_{i-1}/r_i)}{r_i^3 - r_{i-1}^3}, D_F = 3 \end{cases} \qquad [21]$$

Transport Processes

The relationship between parameters of the transport and fractal parameters of soil-pore surface or pore space depends strongly on specifics of the fractal medium model. Such detailed models are rarely available in soil studies, and data on transport are rarely used to infer fractal parameters of soils. Instead, the transport of solutes, colloids, and microorganisms in soils is viewed as a fractal process in which either distances of particle movement or waiting times for

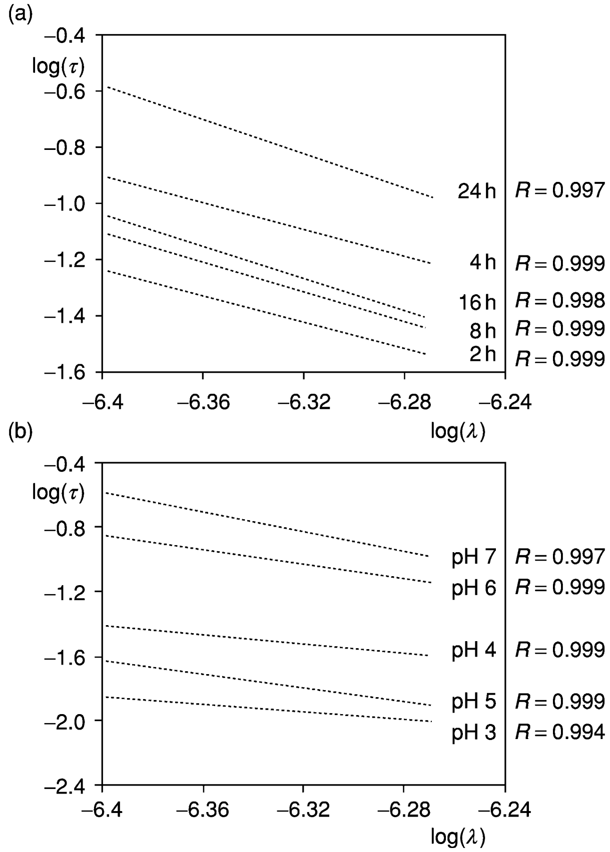


Figure 15 Fractal scaling in dependencies of turbidity τ on wavelength λ in aqueous suspensions of humic acid from Elliot silt loam: (a) at pH 7 after various equilibration times; (b) at various pH values after 24 h equilibration. Fractal dimensions decrease from 2.08 to 1.45 as pH grows from 3 to 7 and morphology of the humic particles becomes less irregular. Reproduced with permission from Senesi N, Rizzi FR, and Acquafredda P (1996) Fractal dimension of humic acids in aqueous suspensions as a function of pH and time. *Soil Science Society of America Journal* 60: 1773–1780.

particles to move have power-law distributions reflecting fractal-like scaling in time or in space. This type of distribution manifests itself in a dependence of dispersivity and diffusivity on the distance from the source of particles participating in transport. Using fractional derivatives and equations of fractional kinetics shows a promise in simulating fractal transport.

Uncertainty in Measuring Fractal Parameters

While it is true that soil is not made up of spheres and straight lines, it is also true that soil is far from being precisely fractal. Rather it is hoped that fractal models may lead to a more accurate description of soil in process models than methods of classic geometry. However, relationships of fractal geometry can be only approximately true in soils. Different methods of characterizing scaling lead to different results even for ideal geometric fractals. Values of fractal parameters obtained by different methods should never be compared for soils. Using proxy measurements creates additional uncertainty, since ideal shapes are assumed to convert nongeometric to geometric measures.

A computational uncertainty also exists in selecting the range of measurement scales in which the fractal scaling is valid. Cutoff values have to be included in the fractal parameter estimation procedure along with slope and intercept of the linear regression used to estimate the fractal dimension and the initiator measure.

The Use of Soil Fractal Parameters

The major contribution that fractal geometry can make to a description of soil is a general, simple, and succinct representation of complex structure by a small number of fractal parameters. Fractal parameters are complementary to other soil parameters. Changes in fractal parameters along with changes in other soil parameters reflect effects of soil management and ecosystem functioning on soil properties. Relationships between structure-dependent different soil properties can be derived from a fractal model when it is selected. Relations between hydraulic conductivity and water retention give one such example. Fractal analysis can help to suggest an iterative process that has caused the observed ‘property-size’ distribution. Cutoffs of the fractal scaling range indicate the change in processes causing the scaling or the change in the performance of a measurement device. They are useful in defining the representative elementary volume or area.

The presence of fractal scaling provides simple ways to simulate soil structure and structure of soil cover. Multiple realizations of such simulations create synthetic data essential to study the effect of soil variability on the performance of new measurement techniques for determining soil properties. It has been found in several research fields that fractal features arise as a result of underlying chaos. Growing interest in the predictability of soil behavior in a changing environment may induce the search for chaos in soils in order to relate it to observed fractal scaling in soil properties. Some time series in soil exhibit scaling properties and can be analyzed using fractal models. Data-intensive sensor technologies such as remote sensing imagery and laser altimetry are becoming more common in soil studies. Fractal-based data analysis techniques can be used to compress these observations into a small number of parameters for comparison and classification purposes.

Fractals should never be considered as an ultimate model of heterogeneity in the soil system. Rather they provide a balance between accuracy and clarity that may aid us in gaining insight into sources and consequences of the observed soil complexity. Eventually, once a greater insight into key processes is obtained, we expect causes of the apparent fractal scaling to be revealed and quantified.

List of Technical Nomenclature

β	Slope of the regression line of wave number versus the power spectrum
γ	Semivariogram
Δ	Grid mesh size (m)
ϵ_{ij}	Residual of the linear regression in the *j*th data point in the *i*th window
λ	Wavelength (m)
σ	Surface tension of the liquid adsorbate ($\mathrm{N\,m^{-2}}$)
τ	Turbidity (m)
A	Pore outline area ($\mathrm{m^2}$)
A_i	Area of the 'condensed adsorbate–vapor' equilibrium interface
A_m	Area of the monomolecular coverage of spherical particles ($\mathrm{m^2}$)
D_F	Fragmentation fractal dimension
D_L	Line fractal dimension
$D_{L,B}$	Kolmogorov dimension
D_m	Mass fractal dimension
D_S	Surface fractal dimension
D_{ss}	Self-similarity dimension
E	Statistical moment about zero
H	Hurst exponent
h	Distance between observation points (m)
I	Intensity of the scattered radiation ($\mathrm{mol\,m^{-2}\,s^{-1}}$)
K	Scaling exponent
M	Number of erosion–dilation operations
m_i	Number of points in the *i*th window
N	Number of pieces into which the object is divided
n	Adsorption ($\mathrm{mol\,kg^{-1}}$)
N_B	Number of squares N_B through which any part of line passes
N_M	Number of visible macropores
N_m	Number of molecules needed for monomolecular coverage of a surface
n_m	Maximum adsorption ($\mathrm{mol\,kg^{-1}}$)
N_r	Number of fragments
n_w	Total number of windows of length w
P	Pore outline perimeter (m)
p	Partial pressure ($\mathrm{N\,m^{-2}}$)
p_0	Partial pressure at saturation ($\mathrm{N\,m^{-2}}$)
Q	Scattering vector ($\mathrm{m^{-1}}$)
q	Order of the statistical moment
R	Reduction factor
R	Gas constant ($\mathrm{J\,mol^{-1}\,K^{-1}}$)
r	Radius of fragment, particle, molecule, pore (m)
r_i	Mean radius of the liquid–vapor interface
RMS	Root-mean-square of residuals
s, S	Scaling exponent
T	Temperature (K)
V_0	Volume of the initiator ($\mathrm{m^3}$)
V_m	Molar volume of the liquid adsorbate ($\mathrm{m^3\,mol^{-1}}$)
W	Scale-dependent measure
w	Window length (m)
x, y	Cartesian coordinates (m)
Z	Observation on a self-affine line

Further Reading

Anderson AN, McBratney AB, and Crawford JW (1998) Applications of fractals to soil studies. *Advances in Agronomy* 63: 1–76.

Avnir D, Farin D, and Pfeifer P (1985) Surface geometric irregularity of particulate materials: the fractal approach. *Journal of Colloid and Interface Science* 103: 112–123.

Baveye P, Parlange J-Y, and Stewart BA (eds) (1998) *Fractals in Soil Science*. Boca Raton, FL: CRC Press.

Falconer KJ (1997) *Techniques in Fractal Geometry*. Chichester, UK: John Wiley.

Giménez D, Perfect E, Rawls WJ, and Pachepsky YA (1997) Fractal models for predicting soil hydraulic properties: a review. *Engineering Geology* 48: 161–183.

Mandelbrot BB (1982) *The Fractal Geometry of Nature*. San Francisco, CA: W. H. Freeman.

Molz FJ and Liu HH (1997) Fractional brownian motion and fractional gaussian noise in subsurface hydrology: a review, presentation of fundamental properties, and extensions. *Water Resources Research* 33: 2273–2286.

Pachepsky YA, Crawford JW, and Rawls WJ (eds) (2000) *Fractals in Soil Science*. Amsterdam, the Netherlands: Elsevier.

Perfect E and Kay BD (1995) Applications of fractals in soil and tillage research: a review. *Soil and Tillage Research* 36: 1–20.

Russ JC (1994) *Fractal Surfaces*. New York: Plenum Press.

Senesi N (1996) Fractals in general soil science and in soil biology and biochemistry. In: Stotzky G and Bollag J-M (eds) *Soil Biochemistry*, vol. 9, pp. 415–472. New York: Marcel Dekker.

Turcotte DL (1996) *Fractals and Chaos in Geology and Geophyisics*. Cambridge, MA: Cambridge University Press.

FREEZING AND THAWING

Contents

Cycles

B Sharratt, USDA Agricultural Research Service, Pullman, WA, USA

Introduction

Soil freezing and thawing are of economic importance and can be an environmental concern in cold regions, because these processes influence erosion, runoff, and greenhouse gas emissions. Soil freezing and thawing are typically confined to latitudes greater than 40° in the Northern and Southern Hemispheres and are largely governed by atmospheric conditions. Soil freezing and thawing are also influenced by surface characteristics such as amount and orientation of crop residues, depth and type of tillage, and snow cover. Soil and crop residue management can therefore influence the occurrence of freezing and thawing. Management practices that result in little residue cover, soil consolidation, or a thin snow pack generally hasten soil freezing in the autumn and soil thawing in the spring. In cold regions, plastic and bituminous mulches are applied to soil to enhance solar radiation absorption and thus expedite soil thawing or prevent soil freezing.

Global Significance

Soil freezing and thawing are confined to latitudes greater than approximately 40° in the Northern and Southern Hemispheres. Soils in tropical and subtropical regions rarely freeze, except at high elevations. The latitudinal zone of ephemeral frost, between the latitudes of approximately 40° and 60°, encompasses regions of seasonally frozen soil. These soils remain frozen for 1 to more than 180 days, go through a single to as many as 100 freeze–thaw cycles, and completely thaw during some portion of the annual cycle. Ephemeral frost occurs in regions typified by continental, marine, and highland climates. Permafrost typically occurs at latitudes greater than 60°; soils at these high latitudes are some of the least productive in the world due to factors such as low fertility and poor profile development. Dry and cold growing seasons also limit crop adaptation and production in permafrost regions. Although seasonal freezing and thawing occur in permafrost (often referred to as the active layer), such events are generally singular during the annual cycle.

Freezing and thawing are important processes that affect the quality and productivity of soil resources. These processes modify soil properties such as thermal conductivity, electrical conductivity, hydraulic conductivity, infiltration, density, shear strength, penetrability, porosity, aggregate stability, and structure. Changes in these properties influence soil water content and availability, nutrient availability, erosion, trafficability of soils, rooting density and distribution, and growth and productivity of agricultural crops. Freezing and thawing also affect the population and diversity of soil microorganisms and fauna. These organisms can be beneficial for enhancing soil aggregation and hastening the decomposition of organic material in soil. Many soil organisms are able to survive subfreezing temperatures by supercooling and entering freeze-resistant stages or by migrating to regions within the soil profile that remain unfrozen.

Water resources are affected by freezing and thawing as a result of redistribution of water within the soil profile or by changes in the rate of infiltration of water into soil. Redistribution of water in the profile occurs in response to water potential gradients that arise as soils freeze. Water migrates from regions of high water potential (a large amount of liquid water in the soil matrix) to regions of low water potential (a small amount of liquid water in the matrix) in the soil profile. This phenomenon is an important mechanism by which recharge occurs in the upper portion of the soil profile due to water migration from wet, unfrozen subsoil to the advancing freezing front. Frozen soil impedes infiltration and thereby enhances runoff. Runoff may carry sediment and nutrients that degrade water resources. Freezing and thawing can also influence the availability and quality of water resources due to the creation of large vertical fractures within the soil profile. These fractures, created by freezing-induced shrinkage stresses, allow for rapid infiltration of surface water and recharge of groundwater. Water flowing through these fractures may be laden with chemicals that contribute to the degradation of groundwater resources in cold regions.

Soil freezing can influence the air quality in cold regions. Freeze-drying of soil during winter results in the breakdown of aggregates into particulate matter; the finest fraction of this particulate matter is subject to transport by wind. Fine particulate matter in the atmosphere, resulting from wind erosion, adversely affects human health. Freezing and thawing also influence greenhouse gas emissions. During winter, emission of these gases is constrained by the relatively impervious frozen soil layer. As soils thaw during spring, however, emission of greenhouse gases is accentuated by either saturated conditions in the top soil (favors nitrous oxide emission) or a burst in microbial activity (favors the emission of carbon dioxide).

Guiding Principles

The frequency and rapidity of soil freezing and thawing are largely governed by atmospheric conditions. Solar radiation and air temperature determine to a great extent the quantity of energy available to modify soil temperatures. Soil freezing is generally of greater risk at locations with lower seasonal insolation or high elevation. In addition, soil surface characteristics also affect the frequency and rapidity of soil freezing and thawing. Surface characteristics such as soil roughness and crop residue cover influence the rate of heat exchange between the soil and atmosphere and thus soil freezing and thawing. Soil and crop residue management is therefore important in determining the frequency and duration of soil freezing and thawing.

Freezing and thawing of soil subject to tillage and residue management practices in cold regions are governed by the heat and water balance of the snow pack, residue layer, and soil (Figure 1). The change in the physical state of the soil is largely driven by atmospheric conditions. Atmospheric conditions such as solar radiation, wind, humidity, air temperature, and precipitation determine to a large extent the amount of energy and water available for heating and evaporative processes at the exchange surface (snow, residue, or soil). Net heat flow must be balanced by heat storage or phase change within a freezing or thawing snow–residue–soil system. In addition, net water flow must be balanced by an increase or decrease in water stored within the system. Heat and water flow processes are intimately related in a freezing or thawing snow–residue–soil system. Within the

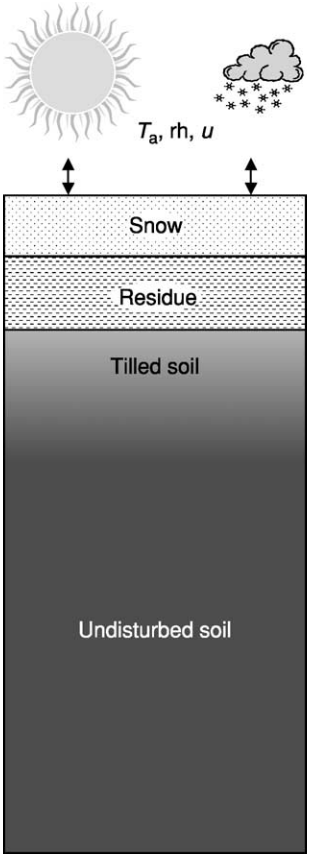


Figure 1 Exchange processes within a snow–residue–soil system governed by solar radiation, air temperature (T_a), relative humidity (rh), wind speed (u), and precipitation.

snow pack, water and water vapor influence heat transfer according to:

$$\delta(k_{\mathrm{sp}}\delta T\delta z^{-1})\delta z^{-1} = C_{\mathrm{i}}\delta T\delta t^{-1} + \rho_{\mathrm{l}}L_{\mathrm{f}}\delta w_{\mathrm{s}}\delta t^{-1} + L_{\mathrm{s}}(\delta\rho_{\mathrm{v}}\delta t^{-1} + \delta q_{\mathrm{v}}\delta z^{-1}) - S \quad [1]$$

where k_{sp} is thermal conductivity of snow, T is temperature, z is depth, C_{i} is volumetric heat capacity of ice, t is time, ρ_{l} is density of liquid water, L_{f} is latent heat of fusion, w_{s} is volumetric liquid water fraction, L_{s} is latent heat of sublimation, ρ_{v} is water vapor density, q_{v} is water vapor flux, and S is an energy source or sink. Equation [1] indicates that any difference in heat flow between the upper and lower boundary of the snow pack (the term on the left side of the equality sign in eqn [1]) must be accounted for by heat storage (first term on the right side of eqn [1]), latent heat associated with phase changes (second, third, and fourth terms on the right side of eqn [1]), or an energy source or sink such as solar and long-wave radiation (fifth term on the right side of eqn [1]).

In the residue layer, heat flow is influenced by water flow according to:

$$\delta(k_{\mathrm{r}}\delta T\delta z^{-1})\delta z^{-1} = C_{\mathrm{r}}\delta T\delta z^{-1} + L_{\mathrm{v}}E - S \quad [2]$$

where k_{r} is the combined conduction–convection heat transfer coefficient for crop residue, C_{r} is volumetric heat capacity of crop residue, L_{v} is latent heat of vaporization, and E is rate of evaporation from the surface of residue elements to void spaces between elements. Equation [2] indicates that any difference in heat flow between the upper and lower boundary of the residue layer (term on the left side of the equality sign in eqn [2]) must be accounted for by heat storage (first term on the right side of eqn [2]), latent heat associated with evaporation (second term on the right side of eqn [2]), or an energy source or sink (third term on the right side of eqn [2]).

In the soil profile, water flow in response to water potential gradients creates an additional complexity in describing heat flow. This complexity arises due to thermal advection that accompanies water flow. Heat flow within a freezing and thawing soil is described by:

$$\delta(k_{\mathrm{s}}\delta T\delta z^{-1})\delta z^{-1} = C_{\mathrm{s}}\delta T\delta t^{-1} - \rho_{\mathrm{i}}L_{\mathrm{f}}\delta\theta_{\mathrm{i}}\delta t^{-1} + L_{\mathrm{v}}(\delta\rho_{\mathrm{v}}\delta t^{-1} + \delta q_{\mathrm{v}}\delta z^{-1}) - C_{\mathrm{l}}\delta(q_{\mathrm{l}}T)\delta z^{-1} - S \quad [3]$$

where k_{s} is thermal conductivity of soil, C_{s} is volumetric heat capacity of soil, ρ_{i} is density of ice, θ_{i} is volumetric ice content, C_{l} is volumetric heat capacity of water, and q_{l} is liquid water flux. Equation [3] indicates that any variation in heat flow within the soil profile (term on the left side of the equality sign in eqn [3]) must be accounted for by heat storage (first term on the right side of eqn [3]), latent heat associated with phase changes (second, third, and fourth terms on the right side of eqn [3]), advection associated with soil water flow (fifth term on the right side of eqn [3]), or an energy source or sink (sixth term on the right side of eqn [3]).

The composition and structure of the uppermost exchange surface will govern heat and water flow processes at the surface and therefore influence the state of the soil system. Soil and crop residue management can modify heat and water flow processes at the upper surface by altering snow cover, residue layer thickness and orientation, or depth of the tilled soil layer. In addition, soil amendments can influence heat and water exchange processes at the soil surface and thus freezing and thawing.

Tillage Practices

Tillage is most often used to prepare seedbeds and to control weed infestations. Tillage can also be used to alter the physical properties at the soil surface to control erosion. In cold regions, tillage practices that hasten soil warming during spring are most desirable. Soil warming can be accomplished by reducing the amount of crop residue on the soil surface or by roughening and darkening the soil surface. Removing crop residue or roughening the soil surface maximizes solar radiation absorption and thus the amount of energy available to warm soil. For more than a quarter of a century, no-tillage has been advocated for reducing wind and water erosion on agricultural lands. This practice, however, retards soil warming during spring due to the insulating effect of crop residue on the soil surface. Therefore, no-tillage is seldom a viable and sustainable option in cold, humid regions.

Tillage practices in common use throughout the world include no-tillage, strip tillage, ridge tillage, moldboard plowing, and various forms of reduced tillage using chisels, sweeps, and disks. The soil microclimate created by any one tillage practice will be quite different from that created by other tillage practices due to the effect of tillage implements on crop residue cover and soil disturbance. Any type of tillage practice results in partial burial of crop residues as well as reduction of soil density within the tilled zone. A reduction in residue cover removes the protective insulation at the soil surface and thereby enhances the likelihood of more frequent and rapid soil freezing and thawing. The rapidity of soil freezing and thawing as influenced by tillage practice is

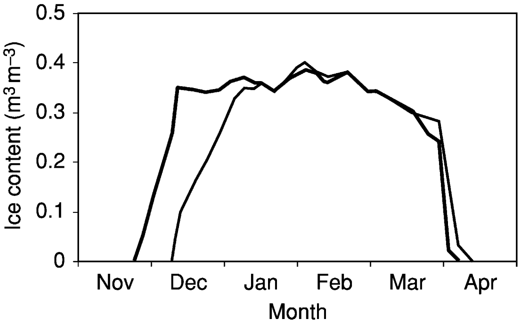


Figure 2 Ice content during winter of a soil subject to autumn chisel plow (thick line) and no tillage. (Adapted from Grant RF, Izaurralde RC, and Chanasyk DS (1995) Soil temperature under different surface managements: testing a simulation model. *Agricultural and Forest Meteorology* 73: 89–113 with permission.)

Table 1 Maximum depth of frost penetration during a winter with little snowfall at three locations

	Frost depth (cm)		
Tillage	*Minnesota*	*Oregon*	*Sweden*
Moldboard plow	74	26	61
No tillage, no residue	76	26	65
No tillage, residue	70	15	51

Sources of information include: Benoit GR, Young RA, and Lindstrom MJ (1988) Freezing-induced field soil water changes during five winters in west central Minnesota. *Transactions of the ASAE* 31: 1108–1114; Pikul JL, Zuzel JF, and Greenwalt RN (1986) Formation of soil frost as influenced by tillage and residue management. *Journal of Soil Water Conservation* 41: 196–199; Thunholm B and Hakansson I (1988) Influence of tillage on frost depth in heavy clay soil. *Swedish Journal of Agricultural Research* 18: 61–65.

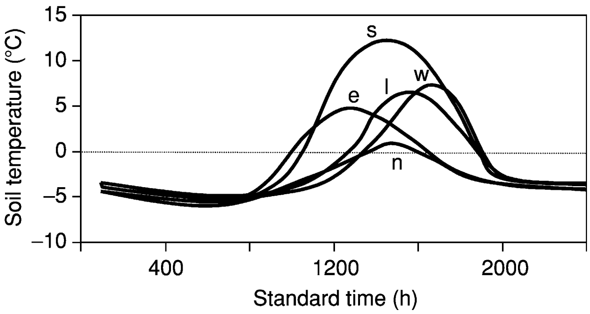


Figure 3 Near-surface soil temperature on various aspects of ridged and level (l) soil surfaces during a clear spring day in Minnesota. (n, north; s, south; e, east; w, west.)

portrayed in Figure 2 by the change in ice content of soil during the autumn and spring. More rapid formation or disappearance of ice is apparent in a soil subject to autumn tillage versus no tillage. Therefore, no-tillage retards cooling of the soil during autumn and warming of the soil during spring as compared to other tillage practices.

Cultivation of soil (either with or without crop residue on the soil surface) reduces the thermal conductivity within the tilled layer and thus the rapidity of heat loss and frost penetration during winter. The maximum depth of frost penetration can be reduced by several centimeters as a result of tilling a bare soil (Table 1). The extent to which tillage influences soil frost penetration, however, will vary across regions and landscapes as a result of the spatial variability in snow cover. Snow cover tends to mask any difference that soil surface characteristics (roughness, residue cover) may have on heat exchange processes. Indeed, as little as 15 cm of snow is sufficient to insulate the soil from atmospheric extremes. Changes in soil thermal conductivity associated with tillage will also influence the rate of soil thaw during spring. In the absence of snow cover, tilled soil will thaw at a slower rate during spring than a bare, nontilled soil. Complete thaw of the soil profile, however, will occur several days earlier in the spring for a tilled versus nontilled soil. Thus, although the rate of thaw is retarded by tillage, reduced frost penetration during winter will result in an earlier thaw of a tilled soil profile versus an untilled soil profile during spring.

Strip-and-ridge tillage systems have been used as a management strategy to modify the soil microclimate. Strip tillage implies row cultivation after harvest, resulting in alternate bands of cultivated soil and residue-covered soil. The width of the band of cultivated soil may vary, but little additional advantage in modifying soil microclimate is achieved for band widths greater than 20 cm. The principal advantage of this strategy is to bolster solar radiation absorption within the cultivated row while retaining residue on the soil surface to control erosion. The additional energy absorbed can expedite thaw by several days and increase temperatures by several degrees within the cultivated rows during spring. Ridge slope and aspect will influence radiation absorption and thus the frequency and duration of soil freezing and thawing. Ridges with a southerly exposure offer the advantage of absorbing more radiant energy than any other ridge aspect in the Northern Hemisphere. Ridges with a south-facing aspect thaw sooner and attain a higher daytime temperature than most other ridge aspects. East-facing aspects thaw earliest on clear days, although daytime temperatures are suppressed during midday as compared to west-facing aspects and level surfaces. Soil thaw and warming are most frequently delayed on northerly aspects (Figure 3).

Crop Residue Management

Crop residue is an effective material that can minimize the erosion of soil by wind and water. Residues reduce the forces of wind and water that would otherwise act upon loose particles at the soil surface. Residues also provide thermal protection to plants from winter temperature extremes. Residues insulate the soil surface from both winter and summer atmospheric extremes by impeding the movement of heat and water vapor between the soil and atmosphere. Crop residue, therefore, retards heat loss from the soil during winter as well as hinders warming of soil during summer.

Crop-residue management includes strategies that alter the amount and orientation of crop residue on the soil surface. The amount and thickness of residue lying on the soil surface determine to a large extent the amount of thermal insulation afforded by the residue layer. For example, an increase in the thickness of the residue layer will enhance the resistance for heat and water vapor exchange and therefore sensible heat transfer between the soil and atmosphere. Percentage residue cover is also important in determining the proportion of solar radiation intercepted by the residue and the soil surface. Radiant energy absorbed by residue is largely dissipated as sensible heat to the atmosphere, thereby resulting in little transfer of heat to the soil. Residue orientation refers to the architecture of the residue canopy; orientation can range from standing stubble to residue lying prostrate on the soil surface. Standing stubble dissipates wind energy at the soil surface and thus minimizes the effectiveness of heat and water vapor transfer by convection from the soil to atmosphere. In cold and windy regions, stubble is important in retaining or trapping snow. Standing stubble also influences the interception of solar radiation; taller stubble generally traps more incident radiation and thus reduces the proportion that is reflected from the residue surface.

Residue management strategies commonly employed to alter soil microclimate or protect the soil from erosion include cutting stubble at various heights above the soil surface at the time of harvest, burning residue in the autumn or spring, and removing residue from the row (row cleaning) at the time of sowing. Perhaps the most effective management strategy for controlling erosion and moderating the soil microclimate has been to cut stubble at various heights above the soil surface. The winter thermal regime of the soil can be dramatically altered by stubble height, particularly in regions where snow is blown by strong winds. Taller stubble retains or traps more snow on the soil surface; the additional snow cover better insulates the soil from winter atmospheric extremes. Taller stubble effectively reduces frost penetration and hastens thawing in the spring (Table 2).

Table 2 Maximum frost depth and day of complete soil thaw during a cold, dry winter in Minnesota

Stubble height (cm)	*Frost depth (cm)*	*Day of thaw*
0, no residue	114	18 May
0	102	14 May
30	75	2 May
60	40	17 April

Adapted from Sharratt BS (2002) Corn stubble height and residue placement in the northern US Corn Belt: I. Soil physical environment during winter. *Soil and Tillage Research* 64: 243–252, with permission from Elsevier.

Burning residue is an effective strategy commonly employed in the autumn before the first snowfall and in the spring after thaw. The ash-covered or bare surface resulting from burning the residue offers little thermal protection to the soil. Soils subject to residue burning, therefore, rapidly freeze and thaw.

Removing residue from the seed row either prior to or during sowing is new technology developed to optimize the soil thermal regime in early spring. Little is known, however, concerning the effect of this strategy on the frequency and depth of soil freezing and thawing. Nevertheless, row cleaning results in a narrow band of bare soil that will thaw and warm more quickly in spring. Daily temperatures can rise 2°C as the width of the soil band increases from 0 to 20 cm. Wider bands of bare soil have no effect on soil temperature. The physical properties of residue also influence heat and water transport between the soil and atmosphere. Little is known, however, about the extent to which soil microclimate can be altered by changing residue properties. Residue color has been exploited in an attempt to alter solar radiation absorption of a soil–residue system. Black residue absorbs approximately 15% more radiant energy than natural-colored residue, but little of this additional energy is utilized in soil thawing or warming.

Soil Amendments

Soil amendments are often used in cold regions to modify the thermal environment of plants. Plastic and bituminous mulches are common amendments used to bolster soil temperatures or expedite snowmelt. These mulches maximize radiation absorption during the day as well as minimize soil heat loss during the night. Plastic mulches not only have the capability to moderate soil temperatures, but they can also modify the soil moisture regime by reducing evaporative loss from the soil. The color, thickness, and composition of the plastic material will influence the amount of solar radiation transmitted to and absorbed by the soil surface. Although these properties (color, thickness, composition) influence radiation absorption, color is perhaps the most widely exploited for promoting soil warming. Black plastic effectively absorbs solar radiation due to its low reflectivity, but little of the absorbed radiant energy is transmitted to the soil beneath the plastic (the energy is dissipated as sensible heat from the surface of the plastic to the atmosphere). White plastic transmits and absorbs little solar radiation owing to its high reflectivity. Clear plastic allows a greater proportion of solar radiation to be transmitted to the soil surface than either black or white plastic; therefore soil temperatures are typically greater under clear plastic than black plastic mulch. Near-surface soil temperatures can be enhanced by as much as 10–20°C using plastic mulch in the early growing season. The largest degree of moderation occurs during the daytime, but temperatures can be raised by several degrees during the nighttime.

Bituminous mulch has also been used to moderate soil temperatures in cold regions. These mulches are frequently used to hasten melting of snow and warming of soil during early spring. The low reflectivity of bituminous mulch results in high absorption of solar radiation. The high thermal conductivity of bituminous material and the good contact of the material with the soil allow for efficient transfer of absorbed radiant energy from the mulch to the soil. Soil temperatures can be bolstered by several degrees using bituminous mulch, thereby delaying the occurrence of soil freezing in autumn and accelerating the occurrence of soil thawing in spring.

List of Technical Nomenclature

θ_i	Volumetric ice content in soil
ρ_i	Density of ice
ρ_l	Density of liquid water
ρ_v	Water vapor density
C_i	Volumetric heat capacity of ice
C_l	Volumetric heat capacity of water
C_r	Volumetric heat capacity of crop residue
C_s	Volumetric heat capacity of soil
E	Rate of evaporation from the surface of residue elements to void spaces between elements
k_r	Combined conduction–convection heat transfer coefficient for crop residue
k_s	Thermal conductivity of soil
k_{sp}	Thermal conductivity of snow
L_f	Latent heat of fusion
L_s	Latent heat of sublimation
L_v	Latent heat of vaporization
q_l	Liquid water flux
q_v	Water vapor flux
S	Energy source or sink
T	Temperature
t	Time
W_s	Volumetric liquid water fraction in snow pack
z	Depth

See also: **Conservation Tillage**; **Crop-Residue Management**; **Freezing and Thawing:** Processes; **Mulches**; **Thermal Properties and Processes**; **Zone Tillage**

Further Reading

Benoit GR, Mostaghimi S, Young RA, and Lindstrom MJ (1986) Tillage–residue effects on snow cover, soil water, temperature and frost. *Transactions of the American Society of Agricultural Engineers* 29: 473–479.

Flerchinger GN and Saxton KE (1989) Simultaneous heat and water model of a freezing snow-residue-soil system. I. Theory and development. *Transactions of the American Society of Agricultural Engineers* 32: 565–571.

Miller RD (1980) Freezing phenomena in soils. In: Hillel D (ed.) *Applications of Soil Physics*, pp. 254–299. New York: Academic Press.

Sharratt BS (2002) Corn stubble height and residue placement in the northern US Corn Belt. I. Soil physical environment during winter. *Soil and Tillage Research* 64: 243–252.

Sharratt BS and Flerchinger GN (1995) Straw color for altering soil temperature and heat flux in the subarctic. *Agronomy Journal* 87: 814–819.

Waggoner PF, Miller PM, and DeRao HC (1960) *Plastic Mulching: Principles and Benefits*. Connecticut Agricultural Experimental Station Bulletin 634. New Haven, CT: Connecticut Agricultural Experiment Station.

Williams PJ and Smith MW (1991) *The Frozen Earth: Fundamentals of Geocryology*. New York, NY: Cambridge University Press.

Wilson C (1969) *Climatology of the Cold Regions Northern Hemisphere. II*. Monograph I-A3b. Hanover, NH: US Army Cold Regions Research and Engineering Laboratory.

Processes

G N Flerchinger, USDA Agricultural Research Service, Boise, ID, USA
G A Lehrsch, USDA Agricultural Research Service, Kimberly, ID, USA
D K McCool, USDA Agricultural Research Service, Pullman, WA, USA

Introduction

Seasonally frozen soil strongly influences runoff and erosion on large areas of land around the world. In many areas, rain or snowmelt on seasonally frozen soil is the single leading cause of severe runoff and erosion events. As soils freeze, ice blocks the soil pores, greatly diminishing the permeability of the soil. This is aggravated by the tendency of water to migrate to the freezing front, causing elevated ice content and frost heave.

Soil freezing and thawing also play a role in a variety of other environmental processes. Frost heave poses significant problems for structures, roads, and plant roots. Soil freezing and thawing can create stress fractures and alter soil physical properties, including pore continuity and aggregate stability; these alterations can influence soil hydraulic properties and erodibility long after the soil is thawed. Water migration associated with soil freezing can strongly influence solute movement. Artificial freezing of the soil has been used to create a barrier in order to isolate contaminants within the soil.

Freezing and thawing of the soil are controlled by the complex interactions of heat and water transfer at the soil surface governed by meteorological and environmental conditions at the soil–atmosphere interface. Different types of frost may form, depending on soil moisture content, rate of freezing, ground cover, and soil physical characteristics. Soil permeability, erodibility, and frost heave depend largely on the type of frost formed.

Heat and Water Relations During Freezing and Thawing

Due to negative water potentials, soil water exists in equilibrium with ice at temperatures below the normal freezing point of bulk water and over the entire range of soil-freezing temperatures normally encountered. When ice is present in the soil, the soil matric potential is strongly influenced by the temperature. As temperature at the freezing front decreases, more and more water freezes, water potential becomes more negative, and liquid water content continues to drop, creating a gradient in water potential and liquid water content. This drop in liquid water content at the freezing front has a similar effect to drying of the soil, and water migrates from moist regions to the freezing front. This often results in elevated ice content, ice lenses, and frost heave.

When ice is present, soil water potential is a function of temperature. This relation is expressed by the Clausius–Clapeyron equation as:

$$\phi = \pi + \psi_{\mathrm{m}} = L_{\mathrm{f}}\left(\frac{T - T_{\mathrm{frz}}}{T}\right) \qquad [1]$$

where ϕ is total water potential, π is soil water osmotic potential, ψ_{m} is soil matric potential, T is absolute temperature, T_{frz} is the freezing point of bulk water (typically 0°C or 273.16 K), and L_{f} is energy, termed the latent heat of fusion, required to freeze water. Thus, when ice is present in the soil, heat and water flux through the soil are tightly coupled, i.e., the matric potential and therefore liquid water content are defined by the temperature and osmotic potential. The relation between matric potential and liquid water content defined by the moisture-release curve is typically assumed valid for frozen conditions.

Darcy's equation can be used to describe steady-state, one-dimensional water flux through the soil:

$$q_{\mathrm{l}} = -K\left(\frac{\partial(\psi_m + \psi_{\mathrm{g}})}{\partial z}\right) \qquad [2]$$

where K is the unsaturated conductivity, ψ_{m} is soil water matric potential, ψ_{g} is gravitational potential, z is depth within the soil, and $\partial(\psi_{\mathrm{m}} + \psi_{\mathrm{g}})/\partial_z$ is the gradient in soil water potential. The transient mass balance equation for water, including the effects of freezing and thawing within the soil, can be written as:

$$\frac{\partial\theta_{\mathrm{l}}}{\partial t} + \frac{\rho_{\mathrm{i}}}{\rho_{\mathrm{l}}}\frac{\partial\theta_{\mathrm{i}}}{\partial t} = -\frac{\partial q_{\mathrm{l}}}{\partial z} + U \qquad [3]$$

Terms on the left-hand side of the equation represent: the time rate of change of liquid water content, and the time rate of change of ice content. Terms on the right-hand side are: the gradient in water flux (i.e., the net flux of water into a layer of soil), and a source/sink term for water. Here, θ_{i} and θ_{l} are volumetric ice and water content, and ρ_{i} and ρ_{l} are the density of ice and water. Eqn [3] states that the net liquid water flux into a soil layer must equal the combined change in ice and water content of the soil. When the net flux is equal to zero, any change in liquid water content must be offset by a change in ice content, adjusted for the difference in density. Although this change in density can result in expansion of the soil matrix, it is not the primary cause of frost heave.

Neglecting vapor flux terms, the equation used for describing one-dimensional heat flow within a partially frozen soil is:

$$C_s \frac{\partial T}{\partial t} - \rho_i L_f \frac{\partial \theta_i}{\partial t} = \frac{\partial}{\partial z}\left[k_s \frac{\partial T}{\partial z}\right] + S \qquad [4]$$

The terms in eqn [4] represent: energy stored in the soil due to a temperature increase; latent heat required to freeze water; net thermal conduction into a layer; and a source term for heat added to the soil. Here, C_s is the volumetric heat capacity of the soil, $\partial T/\partial t$ is the time rate of change of temperature, L_f is the latent heat of fusion, and k_s is soil thermal conductivity. This equation states that the net heat transfer into a volume of soil by thermal conduction and source terms is offset by a change in temperature and a change in ice content of the soil.

Water Migration and Frost Heave

From eqn [1], as the temperature drops below freezing and ice begins to form, the water potential becomes more negative. This creates a gradient in water potential and causes moisture movement toward the freezing zone. Water movement to the freezing zone is described by eqn [2]. If water movement to the freezing zone is sufficient, ice lenses occur, causing the soil matrix to expand. In the case of vertical frost penetration from the soil surface, ice lenses can cause the soil to heave upwards if the pressure associated with freezing exceeds overburden pressures.

The extent of water migration and ice accumulation in the freezing zone is controlled primarily by the rate of freezing front advance in relation to the unsaturated hydraulic conductivity. When the soil is frozen rapidly, there is little opportunity for water to migrate to the freezing front and soil water is essentially frozen in place. Similarly, if the unsaturated conductivity is low, water migration to the freezing front will be slow. Very dry and/or coarse-textured soils have relatively low unsaturated conductivities and exhibit much less frost-related water movement and frost heave than moist, fine-textured soils. Excessive water migration and frost heave are experienced most often when a very moist, fine-textured soil is frozen relatively slowly.

Freezing Dynamics

Soil temperature and water dynamics during soil freezing are shown for a silt loam soil in Figure 1. Continuous measurement of near-surface soil ice content is problematic; however, liquid water content during soil freezing can be measured quite accurately with time domain reflectometry (TDR). Therefore, for illustrative purposes, total water (liquid plus ice) and liquid water content plotted in Figure 1 were simulated by the simultaneous heat and water

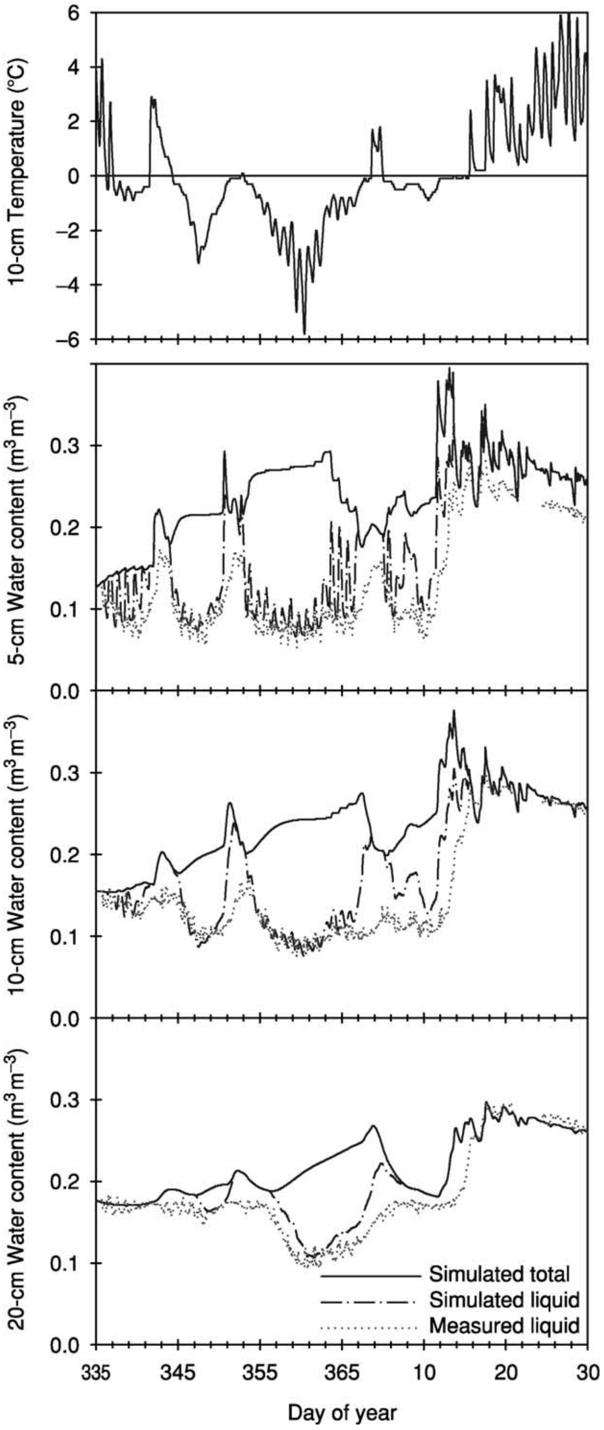


Figure 1 Simulated 10-cm soil temperature, as well as total water content and simulated and measured liquid water content of a silt loam soil for the 5-,10-, and 20-cm depths. (Adapted from Flerchinger GN (2002) Coupled soil heat and water movement. In: *Encyclopedia of Soil Science*. New York: Marcel Dekker, Inc., with permission.)

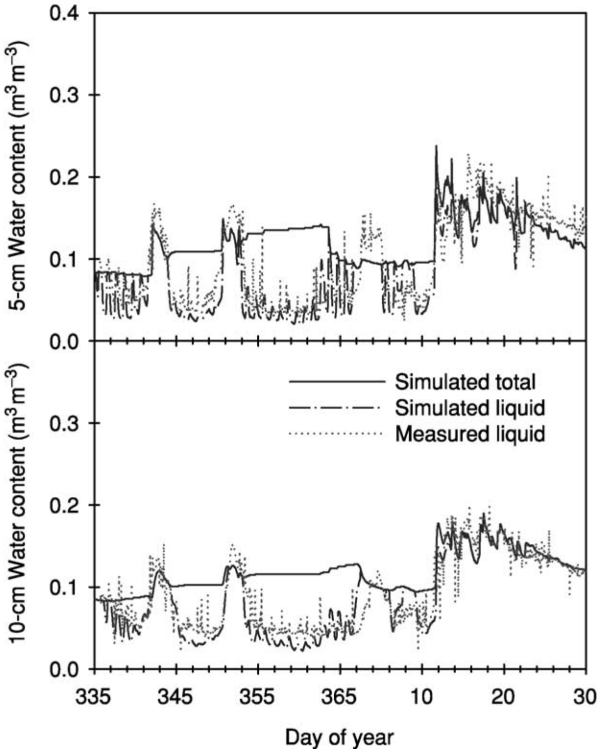


Figure 2 Simulated total water content and simulated and measured liquid water content for a loamy sand soil for the 5- and 10-cm depths. (Adapted from Flerchinger GN (2002) Coupled soil heat and water movement. In: *Encyclopedia of Soil Science*. New York: Marcel Dekker, Inc., with permission.)

(SHAW) model; measurements of liquid water content are plotted for comparison.

The 10-cm soil temperature plotted in Figure 1 shows several freeze–thaw cycles. Frozen conditions can be observed from the water content plotted for the 5-, 10-, and 20-cm depths by separation of the simulated liquid water content line from the total water content line; the difference between the two lines is ice content. Accumulation of ice and an increase in total water content due to water migration to the freezing front can be observed at all three depths plotted. After 9 mm of rain and snowmelt on days 341 and 342, simulated water content above 20 cm was decreasing on day 343 due to drainage. After initiation of soil freezing on day 344, direction of flow reversed, and simulated water flow above 10 cm was upward toward the freezing front. Total water content of the 5-cm depth began to increase on day 344, while liquid water content continued to decrease. As the frost front advanced and the 10-cm depth began to freeze, water migration into the 5-cm soil layer ceased, and the 10-cm total water content began to increase. Subsequently, the 20-cm depth began to freeze on day 347.

Soil water content is shown in Figure 2 for a loamy sand soil, which is a much more coarsely textured soil than the silt loam. Soil water dynamics for the loamy sand are considerably less responsive to freeze–thaw processes than the silt loam soil. Due to the low unsaturated conductivity of the loamy sand, there is much less moisture migration to the freezing front than for the silt loam. As a result, increase in total water content is much smaller.

Freeze–Thaw Impacts on Infiltration

Rain and/or rapid snowmelt on impermeably frozen soil is the leading cause of severe flooding and erosion in many areas of the world. Soil freezing can dramatically reduce the soil's infiltration capacity. Ice blocks the soil pores, resulting in large runoff events from otherwise mild rainfall or snowmelt events.

The permeability of frozen soil is affected by the occurrence, depth, and ice content of the soil, which is dependent on the interrelated processes of heat and water transfer at the soil surface and within the soil profile. Soil freezing and thawing can also alter soil physical properties or structure that impact infiltration. Changes in aggregate stability or stress fractures caused by freezing affect soil structure and pore continuity and thus affect infiltration even after the soil is thawed.

The type of frost formed influences soil permeability after freezing. Soil frost may be divided into four types: granular, honeycomb, stalactite, and concrete. Granular frost is usually found in woodland soils containing organic matter. It consists of small frost crystals, which aggregate around soil particles, but remain separate from each other. Honeycomb frost is commonly found in highly aggregated organic soils and has a loose porous structure, which resembles a honeycomb. Both granular and honeycomb frost typically have high infiltration rates. Stalactite frost often forms in bare soil, which is saturated at the surface. This type of frost consists of loosely fused columnar ice crystals and absorbs water rapidly because of its open porous structure. Concrete frost usually forms in bare, fine-textured, agricultural soils where upward migration of moisture is significant. It is characterized by a complex formation of many thin ice lenses and leaves the ground very hard, much like concrete. Depending on water content, concrete frost can be almost impermeable.

Frozen soil infiltration rates decrease dramatically with soil water content. Water is held less tightly in the large pores, therefore the largest pores that contain water at the time of freezing are the first to freeze. When these larger pores, which conduct water more readily, are blocked by ice, the tortuosity of flow paths increases and permeability and infiltration are severely reduced. Infiltration into a frozen, relatively dry silt loam can typically decrease from 1.0 cm h^{-1} to 0.01 cm h^{-1} as the water content increases to near saturation.

Ice lenses that form due to water migration from the unfrozen subsoil toward the freezing front have an added effect on infiltration. These lenses, typically formed in wet soils having a high proportion of silt, are often a barrier to infiltration. Melting of ice lenses during infiltration influences the temporal variation in infiltration. If ice lenses melt during an infiltration event, infiltration rate can nearly return to unfrozen infiltration rates.

Tillage and surface characteristics that alter porosity and water and heat transport processes can influence infiltration. Tillage processes that create macropores typically have a positive effect on infiltration. Unless the soil is extremely wet, macropores are not filled with water at the time of freezing and remain open for infiltration. However, tillage is likely to have little effect on infiltration if the freezing front descends below the depth of tillage.

Knowledge of frozen-soil infiltration processes lags considerably behind nonfrozen processes. Accurate quantitative descriptions or algorithms of frozen soil infiltration are lacking, partly due to experimental difficulties in measuring infiltration into frozen soil and characterizing the ice content and structure within the frozen soil. With the exception of expensive laboratory techniques such as nuclear magnetic resonance (NMR), there are no quantitative means of directly measuring ice content and structure in the soil, which is the single most important factor affecting infiltration potential upon freezing. Ice content can be computed as the residual between liquid water content measured by TDR and total water content measured by neutron probe or gravimetric samples; however, the sampling volumes of these techniques are dramatically different, making accurate measurement of ice content difficult.

Our understanding of frozen-soil infiltration processes is further hampered by the fact that ice content, pore blockage, and infiltration rate change as water infiltrates into frozen soil. Introducing water into frozen soil causes freezing of the infiltrating water, thawing of the ice contained within the soil, or both. Thus, there is no steady-state infiltration rate analogous to that in unfrozen soil. An approach to circumvent this problem is to use an alternate fluid that remains viscous at subfreezing temperatures. Fluids such as ethylene glycol and air have been used as test fluids for characterizing infiltration of frozen soils. Measured permeability for these alternate fluids can be related to hydraulic conductivities by accounting for differences in density and viscosity.

Various approaches exist for estimating infiltration of frozen soils. Depending on the level of sophistication, adjustments for frozen conditions may be based on: simply whether the soil temperature is below freezing; the amount of ice present in the soil; or the available porosity remaining in the frozen soil. Very simple approaches use essentially a simple on/off switch for accounting for frozen-soil effects, in which the curve number or hydraulic conductivity is set to an arbitrary value to cause reduced infiltration when the soil is frozen. Slightly more sophisticated methods use an adjustment factor to hydraulic conductivity based on antecedent water content or ice content of the soil. Many detailed approaches for estimating infiltration in frozen soils assume the hydraulic conductivity and water retention characteristics are the same for frozen and unfrozen soils. Thus, hydraulic conductivity for infiltration is based on the unsaturated hydraulic conductivity computed from the available porosity (total porosity less volumetric ice content).

Freeze–Thaw Impacts on Soil Erodibility

As soil freezes, water migration to the freezing front can cause ice lenses to form. When the ice lenses melt, the soil often cannot reabsorb all of the excess water, particularly if an impermeably frozen layer still exists below the thawed layer. This supersaturated state results in soil that is extremely weak and susceptible to erosion. However, after drainage and consolidation, soil strength returns. Thus, partially thawed and thawed but unconsolidated soil is highly susceptible to erosion.

Shear strength of a soil is indicative of its resistance to erosion. Specifically, it is defined as the resistance to deformation by the action of tangential (shear) stress. Soil shear strength is made up of cohesion between particles and resistance of particles sliding over each other due to friction or interlocking. Cohesion is composed of true cohesion and apparent cohesion. True cohesion is a function of soil mineralogy and results from chemical bonds between particles. Apparent cohesion, however, is determined by water tension within the soil and is strongly influenced by water content. As the soil thaws at a high water content, soil strength due to apparent cohesion is nil.

Shear stress, τ, caused by water flowing over the soil surface, is defined as:

$$\tau = \gamma R S \qquad [5]$$

where R is the hydraulic radius of the flow, γ is density of water, and S is the slope of the channel of surface. The minimum amount of shear stress required to initiate particle movement is termed the critical shear stress, τ_{cr}, which is a measure of the soil's shear strength. Soil detachment is computed as:

$$D = \alpha(\tau - \tau_{cr})^{\beta} \qquad [6]$$

where α is referred to as soil erodibility and β is a fitted exponent. Studies on thawing soils have shown that moisture content has some effect on τ_{cr}, but the major effect is on α, the soil erodibility. The erodibility of a soil thawed at 5-cm water tension can be approx. 14 times that of the same soil thawed at 45 cm of tension. However, the thawed soil can regain its strength upon drying after tensile forces related to water tension are restored. This strength can be restored in a matter of hours if an impermeable soil layer thaws and allows drainage to occur or if strong evaporative conditions exist. Thus, shear strength can fluctuate dramatically, and timing of rainfall or snowmelt during the first hours of thawing can make a big difference to the erosion that occurs.

Freeze–Thaw Impacts on Aggregate Stability

Aggregate stability, a measure of an aggregate's resistance to breakdown when subjected to external forces, is an important soil property, because soil susceptibility to water and wind erosion increases as aggregate stability decreases, in general. Moreover, many soil physical and hydraulic properties, such as surface-sealing rate, infiltration rate, and hydraulic conductivity, are influenced by aggregate stability. In addition, on medium-textured soils with unstable surface aggregates, crusts can form that hinder or, in some cases, prevent the emergence of seedlings of sown crops.

Mode of Action

As the soil temperature drops below freezing, ice crystals form in the soil matrix, forming first in the soil pores. Once a crystal forms, water flows to the crystal due to a potential gradient, enlarging it, which exerts pressure on nearby aggregates. If those aggregates are constrained and cannot move away from the expanding ice crystal, the pressure exerted upon them can fracture the aggregates directly or develop planes of weakness that can, upon subsequent wet-sieving, cause the aggregates to fracture.

Factors Affecting the Impact

Antecedent water content is the greatest single factor that determines how an aggregate responds to freezing. Aggregate stability decreases, often linearly, with increasing water content at freezing. This relationship appears to hold for many soils ranging in texture from sandy loams to silty clays. As soil water contents increase, more water is available to form ice crystals or ice lenses. Moreover, wetter soil has more water-filled pore space and thicker water films surrounding soil particles, increasing the area through which unsaturated water flow can occur. Also, in wetter soil, tortuosity is less, thereby shortening the water's flow path and speeding its movement to the ice crystal, increasing the latter's rate of expansion. To preserve aggregation and reduce erosion in temperate regions, it is often recommended that producers minimize autumn soil water contents near the soil surface, whenever possible.

The rate at which soil freezes determines in large measure the impact of freezing on soil structure, principally by affecting water redistribution within the soil. If the upper, moist horizons of a soil freeze quickly, that is, if the air temperature decreases sharply, water in those horizons is essentially frozen in place. Consequently, ice lenses, even if they form, do not thicken appreciably and do not compress nearby soil. On the other hand, if a relatively wet soil freezes slowly, water moves to an ice lens and freezes there, thickening the lens, and causing structural deterioration and frost heave.

Soil texture and organic matter also influence aggregate response to freezing. Soils containing high proportions of sand are easily weakened or even fractured. On the other hand, soils with a lot of clay are better able to withstand pressures exerted by nearby ice-crystal enlargement, probably because of additional bond strength provided by more or stronger clay bridges that form between silt and/or sand particles within the aggregate. Organic matter, known to increase the stability of unfrozen aggregates with diameters greater than 0.25 mm, also strengthens aggregates that are later subjected to freezing stresses, providing the water content at freezing is not too great. Elasticity provided by organic matter may enable aggregates from medium-textured (or finer) soils frozen at relatively low water contents to withstand ice-lens expansion pressures before fracturing. Available data indicate that organic matter contents of 3% or more are particularly beneficial for soils that are medium-textured or finer.

The number of freeze–thaw cycles that an aggregate experiences also determines how stable the aggregate is after freezing, with responses being somewhat soil-dependent. In the past, aggregate stability was thought to decrease as freeze–thaw cycles increased, beginning with the first freeze–thaw cycle and continuing monotonically thereafter. Most early research subjected initially air-dried aggregates to many freeze–thaw cycles, often five to ten or more. Recent research, however, has demonstrated that aggregates that have not been air-dried between

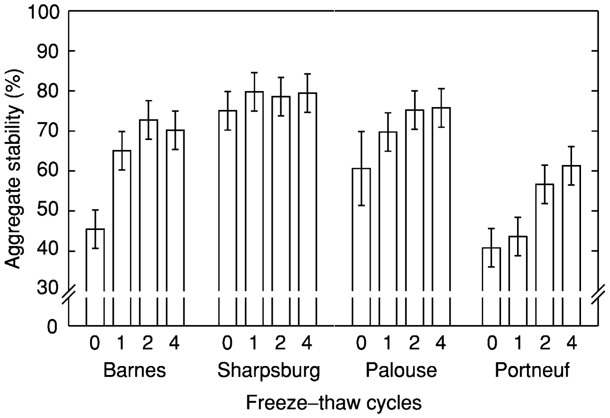


Figure 3 Aggregate stability of four soil types measured after freeze–thaw cycles. (Bars represent 95% confidence intervals.) (Reproduced from Lehrsch GA (1998) Freeze–thaw cycles increase near-surface aggregate stability. *Soil Science* 163: 63–70.)

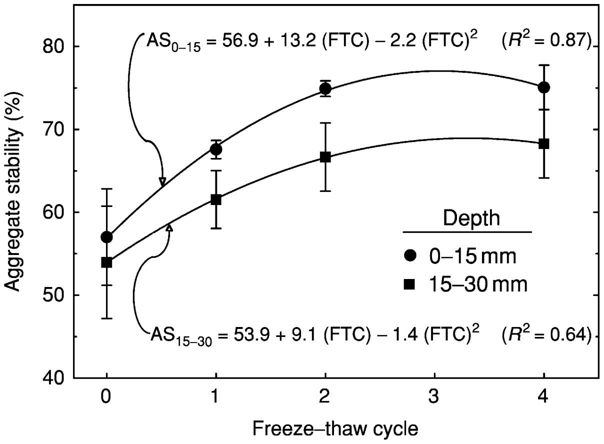


Figure 4 Effects of freeze–thaw cycles on aggregate stability at different soil depths. (Bars represent 95% confidence interval; AS denotes aggregate stability; FTC denotes number of freeze–thaw cycles.) (Reproduced from Lehrsch GA (1998) Freeze–thaw cycles increase near-surface aggregate stability. *Soil Science* 163: 63–70.)

sampling and analysis often increase in stability with the first two or three freeze–thaw cycles (Figure 3). The imposition of many freeze–thaw cycles does indeed decrease aggregate stability; what was seldom recognized was that just a few cycles could increase the stability of aggregates from medium- and fine-textured soils. It has been postulated that ice formation in interaggregate pores and the initiation and early enlargement of ice lenses increase particle-to-particle contacts. Migration of water to the enlarging lens then dries the soil matrix surrounding the ice lens, positioning polysaccharides on soil particle surfaces, gathering and arranging clay domains at points of contact between soil particles, and/or precipitating slightly soluble bonding agents such as $CaCO_3$, silica, or iron oxides at contact points. These processes help aggregates reform and increase in strength, after thawing.

It is particularly interesting that aggregate stability often increases with the first few freeze–thaw cycles but then decreases as more and more freeze–thaw cycles accrue (Figure 4). Opposing forces may be responsible for such phenomena. A force serving to strengthen aggregates may be the result of the precipitation of slightly soluble bonding agents at points of contact between soil particles during the first few cycles. An opposing force that weakens aggregates may be due to ice-lens formation, compression of nearby aggregates, and development of fracture planes. The strengthening process may well occur, and be dominant, for the first two or three freeze–thaw cycles, until most bonding agents have been precipitated from the soil solution. As freeze–thaw cycles continue to accrue, however, more and more fracture planes may be formed in aggregates near where the ice lenses formed. This persistent weakening process may then begin to play the dominant role, decreasing aggregate stability after two or three cycles, as suggested by the fitted curves in Figure 4 and commonly reported in the literature.

Aggregates that are constrained from moving about either in a sample or in a soil profile are weakened more by freezing than are unconstrained aggregates. Aggregates below the soil surface are less stable after freezing than aggregates at or near the soil surface, regardless of the number of freeze–thaw cycles to which they are subjected (Figure 4). Freezing and thawing processes thus affect aggregate stability and soil structure, in addition to water and heat flow through soil.

Soil Freezing and Thawing Effects on Solute Migration

Solutes directly affect water movement during freezing by altering potential gradients. Solutes in the soil solution add an osmotic component to the total water potential. This additional component affects soil water redistribution within the profile as a soil freezes at a given temperature. Specifically, frost heave is greater, that is, more and thicker ice lenses are formed where the soil solute concentration is low, rather than high.

As previously noted, temperature gradients cause soil water to migrate to the freezing front. Solutes in the soil water are carried to the freezing front via convection. As water freezes at the front, solutes are excluded and accumulate in the unfrozen water contained in thin films around soil particles, within soil pores, or in relatively concentrated brine pockets, which may be entrapped within the ice lens itself. If the concentration of a solute in the unfrozen soil solution exceeds its solubility limit, that solute may be precipitated at particle-to-particle contact points, potentially increasing aggregate stability. Upon thawing, the solute transport to the freezing front via mass flow often results in increased concentrations of soluble constituents in the soil water near where the ice lens had formed.

The concentrated brine can also migrate through the frozen soil under certain climatic conditions. When air temperatures at the soil surface are very low, a temperature gradient induces these trapped brine pockets to migrate downward through the ice and frozen soil, into deeper but warmer portions of the soil profile.

See also: **Energy Balance**; **Polar Soils**; **Swelling and Shrinking**

Further Reading

Flerchinger GN and Saxton KE (1989) Simultaneous heat and water model of a freezing snow-residue-soil system. I. Theory and development. *Transactions of the American Society of Agricultural Engineers* 32(2): 565–571.

Formanek GE, McCool DK, and Papendick RI (1984) Freeze–thaw and consolidation effects on strength of a wet silt loam. *Transactions of the American Society of Agricultural Engineers* 27(6): 1749–1752.

Fuchs M, Campbell GS, and Papendick RI (1978) An analysis of sensible and latent heat flow in a partially frozen unsaturated soil. *Soil Science Society of America Journal* 42(3): 379–385.

Iskandar IK, Wright EA, Radke JK, Sharratt BS, Groenevelt PH, and Hinzman LD (eds) (1997) *Proceedings of the International Symposium on Physics, Chemistry, and Ecology of Seasonally Frozen Soils*. CRREL Special Report 97-10. Hanover, NH: US Army Cold Regions Research and Engineering Laboratory.

Kok H and McCool DK (1990) Quantifying freeze/thaw-induced variability of soil strength. *Transactions of the American Society of Agricultural Engineers* 33(2): 501–506.

Lehrsch GA (1998) Freeze–thaw cycles increase near-surface aggregate stability. *Soil Science* 163: 63–70.

Miller RD (1980) Freezing phenomena in soils. In: Hillel D (ed.) *Applications of Soil Physics*, pp. 254–299. New York: Academic Press.

Perfect E, van Loon WKP, Kay BD, and Groenevelt PH (1990) Influence of ice segregation and solutes on soil structural stability. *Canadian Journal of Soil Science* 70: 571–581.

Stahli M, Jansson PE, and Lundin LC (1999) Soil moisture redistribution and infiltration in frozen sandy soils. *Water Resources Research* 35(1): 95–103.

VanKlaveren RW and McCool DK (1998) Erodibility and critical shear of a previously frozen soil. *Transactions of the American Society of Agricultural Engineers* 41(5): 1315–1321.

FUNGI

K Ritz, Cranfield University, Silsoe, MK45 4DT, UK

Introduction

Fungi are ubiquitous in soils, but comprise a varying proportion of the microbial biomass in different systems. They tend to dominate in soils of high organic matter and constitute a smaller proportion in intensively managed mineral soils. They are involved in a plethora of functional roles, encompassing biological, chemical, and physical interactions, and are of great ecological and economic significance. The study of fungi is termed mycology (after *mukēs*, Greek for 'fungus').

Classification

The 'true fungi' belong to the kingdom of Fungi and are eukaryotic, exclusively heterotrophic organisms with cell walls that contain chitin or chitosan as a major constituent, and typically have a thread-like hyphal growth-form, although unicellular forms are common in the yeasts. There are four phyla, the Chytridiomycota, the Zygomycota, the Ascomycota, and the Basidiomycota, plus an informal group denoted the mitosporic fungi (formerly the Fungi Imperfecti or Deuteromycota), which lack a sexual phase and are not ascribed to a formal taxonomic position. Only the Chytridiomycetes produce motile zoospores. Representatives from all taxonomic groupings are commonly found in soils, where fungal biodiversity, as for all soil microbes, is typically several orders of magnitude more than aboveground or aqueous systems.

In an informal sense, organisms studied by mycologists cross the eukaryotic kingdoms of Protozoa and Chromista as well as the Fungi. Within the Protozoa, the cellular slime molds (phyla Dictosteliomycota, Acrasiomycota) and plasmodial slime molds (phyla Myxomycota, Plasmodiophoromycota) form fruiting bodies with a resemblance to some fungal types; the ameboid phagocytotic cells of these organisms are common in soils and are implicated in nutrient cycling. Within the Chromista, the Oomycetes form mycelia but contain some cellulose in their walls and produce motile zoospores. This group includes some important soil-borne plant pathogens within the order Pythiales, such as *Phytopthora infestans* (potato blight) and *Pythium* spp. (damping off).

Biology

Morphology and Form

A fundamental characteristic of the filamentous (eucarpic) fungi is their growth-form, which consists of a thread-like hypha which grows by apical extension and periodic branching to form a mycelium that permeates the environment in which the fungus is growing (**Figure 1**). The diameters of individual hyphae vary according to age, species, and nutritional conditions but are typically 3–10 μm. In extreme cases, up to 2 km of hyphae per gram of soil have been measured in some soils; a typical cultivated arable soil contains several meters per gram. Mycelia are indeterminate structures that grow in three dimensions and therefore conceptually are difficult to size. In soil they may occupy almost any volume, and the largest organisms on the planet are probably mycelia of *Armillaria bulbosa*, individual clones of which are reported to extend continuously over tens of hectares in North American forests (**Figure 1g**). Hyphae are generally compartmentalized by cross-walls (septa) which occur at varying intervals depending on species (**Figure 1b**). Septa may be complete or perforated to varying degrees and play a role in regulating the passage of materials, including organelles, between adjacent compartments. If mycelia are fractured, septal pores become occluded, which prevents leakage of cytoplasm from the open ends of broken hyphae. Hyphae may aggregate to form higher-order structures such as strands (cords; **Figure 1e**) or rhizomorphs, which enable long-distance foraging (decameters in the case of woodland basidiomycetes). Hyphal aggregation and differentiation can produce morphologically complex macrostructures such as the intricate fruiting bodies (basidiocarps) of some Basidiomycetes. Those fungi that do not form hyphae (predominantly yeasts) have a holocarpic growth form, comprising individual cells that replicate by budding or binary fission. Some fungi are dimorphic and can switch between holocarpic and eucarpic growth forms depending on environmental conditions (e.g., *Mucor hiemalis*).

Physiology and Reproduction

All fungi are obligate heterotrophs, i.e., they utilize fixed (organic) C sources as substrate. Across the kingdom an extremely wide range of compounds can be utilized, ranging from C_1 compounds (methylotrophic yeasts) through to lignin and industrially produced polyphenols – almost any naturally occurring macromolecule occurring in soil can be degraded by some fungi. A concomitantly wide range of enzymes and metabolic capabilities are expressed. The range of compounds utilizable by any particular species is usually reflected by their normal environmental niche, and different types vary widely in their capacity to utilize different substrates. Respiration can be aerobic or anaerobic, and obligate and facultative types of both forms occur. Some fungi obtain their nutrition from other living organisms via parasitic or mutualistic associations, whilst others are saprotrophs. Many soil fungi are very efficient scavengers of nutrients and grow sparsely but effectively under the oligotrophic conditions that prevail in many soils. A wide variety of secondary metabolites are produced by soil fungi, including many volatile organic compounds (VOCs) which may act as signals or substrates as they diffuse through pore networks. If fungal spores or mycelial fragments are added to nonsterile soils, they rarely germinate. This phenomenon, known as fungistasis (or mycostasis) is probably due to a lack of available nutrients (the extant microflora normally having depleted the microbially available nutrient pool), the presence of inhibitory (antibiotic) compounds produced by the existing microflora or a combination of both.

Reproduction can be sexual or asexual, and a variety of mating systems occur. Many fungi have genetic systems that prevent mating between genetically identical cells or mycelia. If hyphae encounter each other, they may fuse (anastomose). If they are somatically compatible, the fusion results in the formation of larger mycelial networks, with attendant functional consequences in terms of spatial exploration, activity, translocation, and combativeness. Somatic incompatibility, however, results in a rejection response,

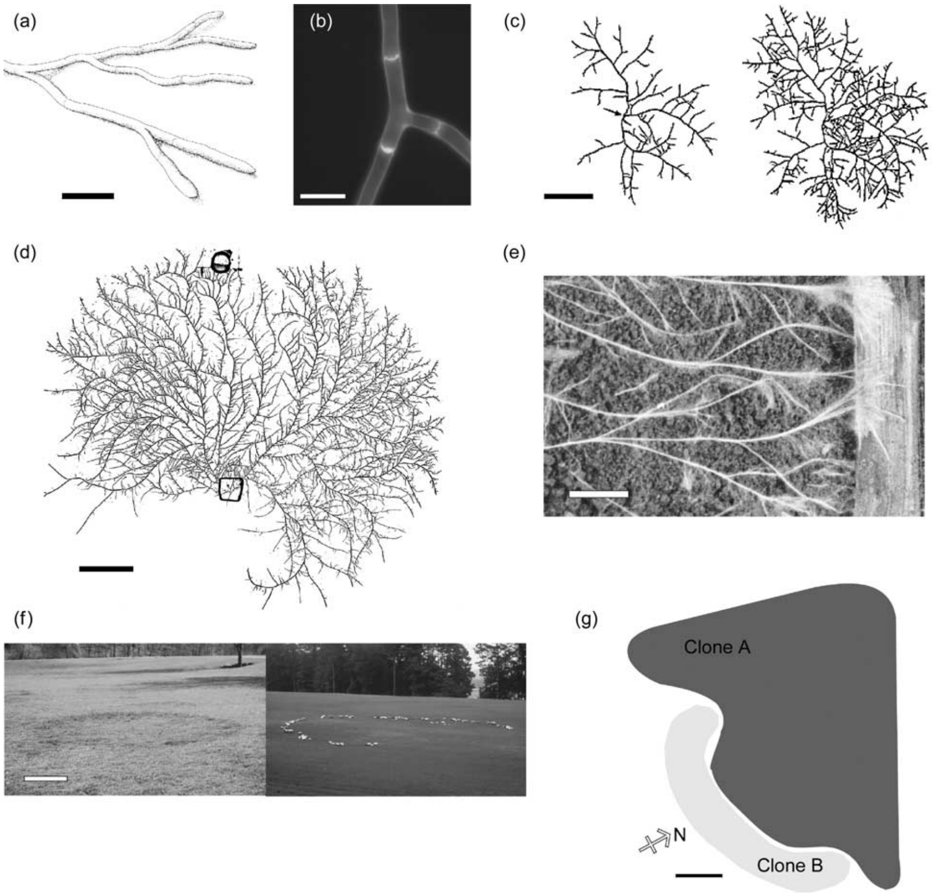


Figure 1 Aspects of mycelia of soil fungi at different scales. (a) stylization of fungal hyphae, showing branching habit and morphology of tips. Scale bar 50 μm; (b) micrograph of hyphae of *Rhizoctonia solani* showing branch and septa. Scale bar 50 μm; (c) mycelium of *Trichoderma viride* 44 h (left) and 48 h (right) after germination, grown on spatially uniform nutrients. Scale bar 100 μm; (d) mycelium of *T. viride* grown on spatially nonuniform nutrients (point of inoculation is denoted by a square, localized nutrient source denoted by a circle). Scale bar 5 mm; (e) colonizing fans at tips of foraging cords of *Agrocybe gibberosa* when a piece of straw (to right of image) is encountered. Scale bar 5 mm; (f) fairy ring in grassland, resulting from enhanced mineralization of soil nutrients by subterranean mycelium, expressed as enhanced plant growth (left image) and following basidiocarp emergence (right image). Scale bar 1 m; (g) mapped extent of two clones of *Armillaria bulbosa* in a Michigan forest. Scale bar 100 m. ((a) is adapted from Jennings DH and Lysek G (1999) *Fungal Biology*. Oxford, UK: Bios Scientific, with permission. (b) is courtesy of Karl Ritz. (c) is reproduced with permission from Ritz K and Crawford JW (1990) Quantification of the fractal nature of colonies of *Trichoderma viride. Mycological Research* 94: 1138–1142. (d) is adapted from Crawford JW *et al.* (1993) Quantification of fungal morphology, gaseous transport and microbial dynamics in soil: an integrated framework utilising fractal geometry. *Geoderma* 56: 157–172, with permission. (e) is reproduced with permission from Robinson CH *et al.* (1993) Resource capture by interacting fungal colonies on straw. *Mycological Research* 97: 547–558. (f) is reproduced with permission from Eric Nelson, Cornell University. (g) is adapted from Smith ML *et al.* (1992) The fungus *Armillaria* bulbosa is among the largest and oldest living organisms. Nature 356: 428–431 © 1992 Macmillan Magazines, with permission.)

and the cytoplasm in fused compartments is destroyed. Fungal propagules generally take the form of asexual or sexual spores, sclerotia (dense, heavily pigmented hyphal aggregations), or hyphal fragments. Some species sporulate very profusely (e.g., *Penicillium*), others have never been known to produce any form of spore (e.g., *Rhizoctonia solani*). In soil, propagules are dispersed by mass flow in soil water, carried by animals internally or externally to their bodies, or via physical (often faunally mediated) soil movement.

Biotic Interactions

In the soil environment, no organism grows in isolation, but always in the context of a wider microbial community; organisms have evolved under such circumstances over millennia. It is not surprising therefore that there are extensive interactions between soil fungi and all other classes of organism, including other fungi. At present we only have a limited understanding of the degree and nature of such interactions, since the majority of microbiological research has been on the growth and physiology of microbes grown in isolation, in simple combination, and often *in vitro*.

Fungal Associations with Other Microbes

Fungal associations with algae or cyanobacteria, in the form of lichens, play a role in soil genesis via rock weathering. In soils proper, when fungal mycelia meet, there can be a variety of outcomes from such interactions. If they are somatically compatible, they will fuse to form a larger mycelium (see above); other outcomes include an intermingling without fusion, chemically mediated repellence, physical exclusion by the development of insulated barriers, or death and replacement of one mycelium by the other. There are complex hierarchies of combativity between different species and strains, further dependent on environmental factors such as temperature, water, and nutritional status. Fungal communities in soil are therefore in a state of perpetual dynamics and interaction. There are also many examples of specifically hyperparasitic fungi (e.g., *Arthrobotrys oligospora*, *Coniothyrum minitans*) and bacterial mycoparasites (e.g., *Pseudomonas* spp.). Mycoviruses do not have an extracellular phase, and since they do not necessarily affect the fungal phenotype they may be more prevalent than was hitherto thought.

Fungal Associations with Plants

Plants are the primary autotrophs in terrestrial systems and therefore responsible for the deposition of fixed C into soils, which provides the majority of substrate for the heterotrophic soil biomass. Biotic associations between vascular plants and fungi are extensive.

Mutualism Mutualistic associations between plant roots and fungi are termed 'mycorrhizas' (Figure 2a–e) and first evolved almost simultaneously with the emergence of land plants. Such associations are extremely common, and the natural state for the majority of root systems is to be infected to some extent by mutualistic fungi. There are four major types of mycorrhizal association that differ anatomically and physiologically and have variable host ranges (Table 1). In the majority of cases, the fungal partner derives carbon from the host, and the plant benefits from an enhanced uptake of nutrients such as P or N and water, mediated by extensive external mycelia. Mycorrhizal infection may also confer resistance against root pathogens. Given that most soil microbes are C-limited, it is likely that mycorrhizal hyphae constitute a significant proportion of the fungal biomass in vegetated soils, since they are intimately associated with a readily available source of C via their host plants. Mycorrhizas strongly affect the competitive ability of plants (often mediated via nutrient acquisition) and thus play a role in governing the community structure of vegetation. Mycorrhizal networks can infect a number of plants of the same or different species simultaneously and thus form an interconnected network, providing the possibility for nutrient transfer between plants via interconnecting hyphae. The ecological significance of such phenomena is unclear.

Parasitism Plant pathogenic fungi cause very significant economic losses to crops in both temperate and tropical agricultural and forestry systems. They also occur in natural systems and play a role in mediating plant community structure, although relatively little is known about the extent of such regulation.

Soil-borne plant pathogens occur across all fungal taxonomic groups and vary in their host specificity; important examples include *Gaeumannomyces graminis* (take-all of wheat; Figure 2f), *Rhizoctonia solani* (root and stem rots, including potato tubers, with a wide host range), *Pythium* spp. (damping-off), *Armillaria mellea* (honey fungus, with a wide host range of trees), *Fusarium oxysporum* (a wide variety of forms, which affect specific plants – e.g., *F. cubense* causes wilt in banana).

The primary site of infection is normally roots, although some pathogens attack the hypocotyls or stem bases of plants on or near the soil surface, and there are a range of tuber diseases. Many pathogens show strong taxes toward roots, mediated via soluble or volatile exudates, or electrical charges (electrotaxis). Oomycete zoospores show negative geotaxis and a tendency to swim against the direction of water flow, which keeps them in upper soil layers where roots are more abundant, as well as electrotaxis toward roots and wound sites. The pathogenic mode of action takes many forms between different pathogens, from necrotrophic attack and wholesale degradation of tissue, through cortical invasion, root hypertrophy, to the blocking of vascular tissue and hence interference with water and solute transport (wilts). There is often a proliferation of mycelia on root surfaces prior to penetration via localized aggregations of hyphae (infection cushions).

Control of root pathogens is often based upon use of resistant cultivars, crop rotation to mitigate build-up of pathogen populations, and appropriate soil management such as tillage practice, organic matter incorporation, or liming. Persistent growth of the same cereal crop can result in so-called take-all decline, where disease severity is reduced after several seasons, thought to be due to the build-up of microbial communities that are antagonistic to *Gaeumannomyces*. Chemical control is rarely feasible due to

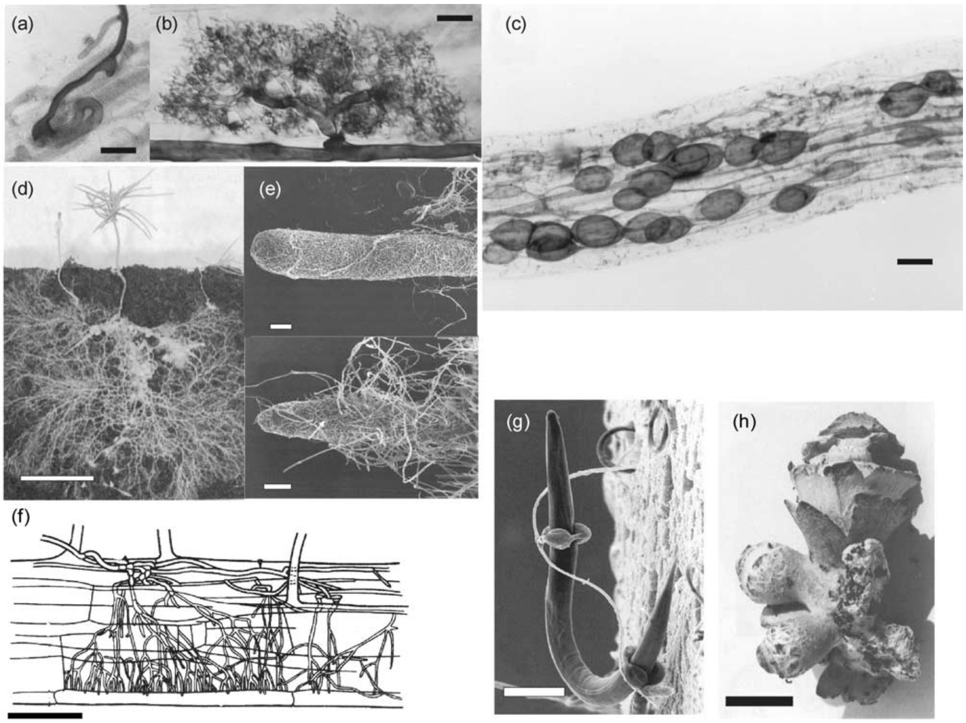


Figure 2 Mutualistic associations involving soil fungi. (a) infection point of arbuscular mycorrhizal (AM) fungus on root of *Plantago lanceolata* (plantain). Scale bar 50 μm; (b) arbuscule in root cell of *Allium porrum* (leek). Scale bar 3 μm; (c) intraradical mycelium and vesicles of AM fungus proliferating in root cortex of *P. lanceolata*. Scale bar 200 μm; (d) external mycelium of *Suillus bovinus* on *Pinus sylvestris* (pine) in microcosm system. Scale bar 5 cm; (e) ectomycorrhizal colonization of roots of *Betula alleghaniensis* (yellow birch) by *Pisolithus tinctorius*. Top: thin mantle of mycelium formed on root surface; bottom: compact mantle entirely enveloping roots, and development of external mycelium. Scale bar 100 μm; (f) infection of *Triticum aestivum* (wheat) roots by *Gauemannomyces graminis* (take-all). Scale bar 250 μm; (g) nematode trapped by constricting rings of *Arthrobotrys anchonia*. Scale bar 100 μm; (h) basidiomata of *Leucoagaricus gongylophorus* removed from nest of *Atta cephalotes* (ant). Scale bar 2 cm. ((a, c) Reproduced with permission from Karl Ritz. (b) Reproduced with permission from Brundrett MC *et al.* (1983) A new method for observing the morphology of vesicular-arbuscular mycorrhizae. *Canadian Journal of Botany* 62: 2128–2134. (d) Reproduced with permission from Harley JL, Smith S, and Reid DJ (1996) Mycorrhizal Symbiosis, 2nd Edition, with permission from Academic Press. (e) Reproduced with permission from Massicotte HB *et al.* (1990). Structure and ontogeny of *Betula alleghaniensis-Pisolithus tinctorius ectomycorrhizae. Canadian Journal of Botany* 68: 579–593. (f) Adapted from Skou JP (1981) Morphology and cytology of the infection process. In: Asher MJC and Skipton PJ (eds) *The Biology and Control of Take-all*. London, Academic Press, with permission. (g) Reproduced with permission from George Barron. (h) Reproduced with permission from Fisher PJ *et al.* (1994) *Leucoagaricus basidiomata* from a live nest of leaf-cutting ant *Atta cephalotes. Mycological Research* 98: 884–888.)

the buffering capacity of soil against bioactive compounds and the rapid degradation of such compounds by the soil microflora.

Fungal Associations with Fauna

Fungal associations with soil fauna are extremely common, including a wide range of mutualistic and parasitic relationships. For example, predatory, nematode-trapping species (e.g., *Arthrobotrys* and *Dactylella* spp.) have evolved an extensive variety of mechanisms to procure their prey, including sticky hyphae, branches, nets, and knobs, which are sufficiently adhesive to prevent escape of nematodes prior to digestion by the fungus (analogous to the insectivorous sundew plants), and hyphal rings (lassoes), some types of which contain inflating trigger cells, which grip the bodies of prey to prevent escape (Figure 2g). Other fungi (e.g., *Verticillium* spp.) infect nematodes via spore attachment and hyphal penetration of cysts, juveniles, or adults. Virtually all insects studied to date appear to have a range of parasitic and mutualistic (principally in their guts) fungi associated with them, including during larval and pupal stages.

Fungal hyphae and spores are also attacked or grazed by representatives from most faunal groups, including protozoa, nematodes, worms, insects, and mites. Subtle interactions occur. For example, there is a wide variation in the degree of palatability of different fungi to different fauna, which results in preferential grazing of mycelia by soil fauna, analogous to aboveground herbivory. Soil fauna play a strong role in dispersing fungi through soil, either as spores attached to their bodies or via egestion in faeces at sites remote from where the fungus was

Table 1 Basic characteristics of the four mycorrhizal types

Mycorrhizal type	*Habit*	*Host range*	*Examples*	*Notes*
Arbuscular (AM)	Endotrophic; form arbuscules and sometimes vesicles in root cortices	*c.* 90% of vascular plants; exceptions include members of Brassicae	Exclusively order Glomales within Zygomycetes (e.g., *Acaulospora, Gigaspora, Glomus*)	Obligate mutualists; can only be cultured in association with plants (including root-organ tissue culture systems)
Ecto- (ECM)	Ectotrophic; form extensive mycelial sheath around root	Mainly woody plants	Basidiomycetes (e.g., *Amanita, Suillus*); Ascomycetes (e.g., *Melanospora*); Zygomycetes (*Endogone*)	May also grow saprotrophically in absence of host plant, and in pure culture
Ericaceous	Endotrophic; extensive intracellular coils formed inside host cortex	Ericaceae	*Hymenoscyphus, Oidiodendrum*	May enhance availability of N to host by degradation of soil organic N
Orchidaceous	Endotrophic; short-lived intracellular coils in host root cells	Orchidaceae	*Rhizoctonia, Marasmius*	Often obligate mutualism required for host plant, to enable seedling development; achlorophyllous orchids are solely dependent on fungus for substrate

ingested. Some Basidiomycetes (e.g., *Termitomyces*, *Acromyrmex*, *Leucoagaricus* spp.; Figure 2h) are 'farmed' by termites and ants, maintained within the insect colonies in essentially pure culture on imported plant material.

Fungal Contributions to Soil-System Function

Nutrient Cycling

Soil fungi play a crucial role in nutrient cycling in terrestrial systems, due to the primary role they play as decomposers, mediated by a large repertoire of catabolic activities. They are particularly involved in cycling C, N, and P, but have roles in most of the other soil elemental cycles. Fungi are primarily involved in the decomposition of plant material, the lignocellulosic components of which are relatively recalcitrant to bacteria. There is typically a succession of fungal (and other microbial) species on new substrate when it enters soil, based upon the initial decomposition of simpler compounds and a subsequent degradation of more complex polymers. Cellulases are produced by a wide variety of fungi across the taxonomic classes; these include the brown rots, so-called because their activity results in a darkening of wood as the cellulose is decomposed and lignin (which they are not able to decompose) remains. Within fungi, polyphenol oxidases (mainly laccases), involved in the decomposition of lignin, are largely confined to Basidiomycetes and Ascomycetes. These include the white rots, species which normally also have strong cellulase activity and render wood a light color during decay.

As well as invoking the enzymatically mediated degradation of polymers, many fungi also produce a variety of compounds such as organic acids or siderophores, which solubilize or immobilize essential or toxic metals, with consequences for both their own growth and the growth of proximal organisms. Fungi also act to immobilize nutrients in soil through sequestration in mycelia, which are further released following death and lysis or attack by pathogens or grazers. Eucarpic fungi play a significant role in physically transporting nutrients through the soil fabric via translocation within mycelia, thus short-circuiting other diffusion and mass-flow transport mechanisms, both in saprotrophic and mutualistic (mycorrhizal) contexts.

Soil Architecture

Eucarpic fungi influence soil structural dynamics via a number of mechanisms and on different spatial scales. Hyphal networks enmesh soil particles together, rather like the bars in reinforced concrete. Sparse mycelia in nutrient-poor mineral soils are unlikely to be particularly influential in this respect, but the dense mycelia that can form where substrate is abundant, for example in forest litter layers, can be quite resistant to physical disruption. These mechanisms operate on scales of millimeters to centimeters. The variety of compounds exuded by hyphae as they grow bind soil particles together via adhesive or charge-based mechanisms, on scales from nanometers (charge) to millimeters (adhesive). Such binding results in increased aggregation of soil and an associated structural generation. Degradative enzymes prevalent near hyphal tips may also decompose organic matter binding soil particles together, resulting in structural degeneration. Fungi produce large quantities of a diverse array of hydrophobic proteins (hydrophobins), which serve to insulate hyphal walls. This is important due to the large surface area-to-volume ratio that arises from the mycelial growth-form. However, these compounds also coat soil particles and can strongly influence the water sorptivity and repellency characteristics of the soil. This in turn affects the hydraulic properties on larger scales. It has recently been suggested that glomalin, a glycoprotein produced in abundance by arbuscular mycorrhizal (AM) fungi and which is resistant to decomposition, may play a key role in effecting mechanisms of adhesion and soil stabilization. As hyphae extend through soil, they may also cause physical restructuring of soil, by rearranging particles, and orientate clay lamellae parallel to their walls (Figure 3).

Whilst fungi certainly affect soil structural dynamics, the physical architecture of the soil also affects fungal growth in that it provides the physical framework through which fungi must grow in order to locate resources. The eucarpic growth-form is particularly effective for growth in such spatially structured environments, where complex pore networks prevail and water and substrate are distributed in a heterogeneous manner within the three-dimensional matrix. Hyphae are able to grow across the surfaces of pore walls but also bridge wider pores; since mycelia form an interconnected network, materials can be transported within them from zone to zone, potentially supplementing growth factors where they are in demand in particular regions. Translocation within most soil fungi is as yet poorly understood. There is evidence that different materials can be simultaneously transported in opposite directions within mycelia. Most substrate in soils is distributed patchily, and path lengths between resource depots can be extended greatly by the tortuosity of pore networks. Many species show foraging strategies which balance explorative and exploitative growth strategies relative to the distribution of resources encountered in their normal habitats. Such strategies are highly developed in the cord-forming Basidiomycetes. Physical exclusion mechanisms, where pore diameters are too small to permit penetration by hyphae, probably operate; the extent to which hyphae can physically separate soil particles to allow penetration, akin to plant roots, is largely unknown. Physical protection of organic matter and organisms (including hyphae) from degradation or predation via such spatial exclusion also occurs. Hyphae are generally poor at growing through water-filled pores and hence are less prevalent in waterlogged soils. There are thus strong interactions between soil structure, water, and mycelial growth.

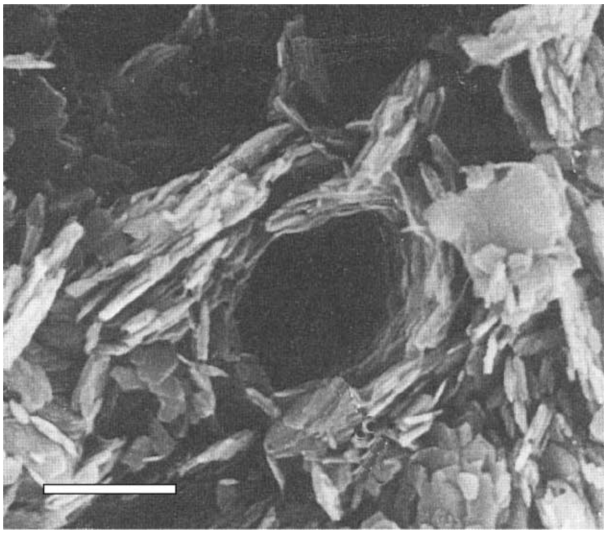


Figure 3 Orientation of kaolinite clay platelets by hyphae of *Chetomium* sp. Scale bar 2 μm. (Reproduced with permission from Dorioz JM *et al.* (1993) The role of roots, fungi and bacteria on clay particle organization. An experimental approach. *Geoderma* 56: 179–194.)

Methods of Study

A comprehensive range of methods has been developed and applied to study the presence, abundance, activity, and distribution of fungi in soils. ‘Quantification’ of soil fungi presents a particular challenge, since there is a diversity of functionally important components, ranging from total hyphal length through mass of cytoplasm and number of propagules. Methods must be carefully matched to the nature of the study being undertaken.

Isolation Methods

A wide diversity of fungi can be isolated from soils by traditional methods of dilution plating and enrichment growth on broad-scale or specific gel-based media supplemented with a variety of nutrient sources, ranging from oligotrophic (based on silica gels, which are devoid of organic compounds) through to C-rich media such as Czapek-Dox medium (an agar gel based on 3% sucrose). Since relatively low dilutions of soil are often necessary compared with those appropriate for bacteria (typically 10^{-3}

in contrast to 10^{-5}), media may be supplemented with antibacterial compounds to inhibit contaminating overgrowth. Many soil prokaryotes are apparently not cultivable *in vitro*, but the extent to which the 'noncultivability syndrome' may apply to soil fungi is as yet unclear. Whilst such plating methods may be appropriate to determine fungal richness (i.e., a basic range of cultivable species present), meaningful interpretation of 'colony-forming units' (CFUs) by such means is often dubious. This is because colonies can form from hyphal fragments as well as spores, and the number of CFUs will therefore be strongly affected by the propensity for, and degree of fragmentation of, mycelia during sample preparation – it is difficult to relate this to the probable status of any mycelia in the soil. In addition, the expression of colonies on media can be dependent on the presence or absence of other species within the same plate. Hyphal fragments or (larger) spores can be directly extracted from soils by micromanipulation, on density gradients or by elutriation and sieving. 'Baiting' methods are based on the insertion of sterile substrates such as gel-filled perforated tubes, cotton strips, or hair into undisturbed soil, followed by removal and isolation of fungi colonizing these substrata. If an air gap surrounds the baits, it is assumed that any fungi entering them were in an active hyphal state in the soil.

Direct Observation

Measurement of total hyphal length can be made by direct observation of thin films of dispersed soil using a microscope. It is normally necessary to stain hyphae using colored (e.g., aniline blue, acid fuchsin) or fluorescent (e.g., Mg ANS, an 8-anilino-1-naphthalene sulfonic acid salt; Fluorescent Brighteners – FB24, also known as Calcofluor W M2R) stains to render them visible against the complex background of soil particles. Metabolically active hyphae can be distinguished from inactive forms by use of vital stains such as fluorescein diacetate (FDA) or tetrazolium dyes. By measuring hyphal diameters as well, estimates of biovolume and thence biomass can be made using conversion factors. Soil films can be produced by smearing small volumes of suspension on to glass slides or pulling suspensions down on to membrane filters. Soil thin-sections, produced by impregnation of undisturbed soil with hard-curing resin, and cutting and polishing the resultant blocks with lapidary (jeweller's) pastes, enable the *in situ* observation of fungal features (Figure 4). Computerized image processing and analysis can be used to automate identification and measurement procedures, but such methods rely on a high degree of contrast between hyphae and background being obtained. It is rarely possible to identify fungi by these methods, since most hyphae do not have a sufficiently distinct morphology. Infection of roots by AM fungi is normally assessed by direct microscopic observation of roots cleared of cytoplasm by treatment with alkali, and stained using dyes such as trypan blue or chlorozol black E.

Molecular Methods

Complementary to cultivation and direct-observation techniques are so-called molecular analyses, which focus on biochemicals that are more or less specific to fungi. These methods circumvent many of the problems associated with *in vitro* cultivation of fungi and offer a wide range of scales of resolution in quantitative analysis. Total fungal biomass in soils can be estimated by direct extraction and measurement of biomarker compounds such as chitin or ergosterol. The effectiveness of this approach relies on such compounds being specific to fungi and present in a constant proportion relative to the amount of biomass. Few compounds meet these criteria strictly. For example, chitin is also present in arthropods and ergosterol concentrations can vary significantly depending on the age, nutritional status, and species of fungus. More potential is shown by phospholipid fatty acids (PLFAs); these membrane-bound compounds vary in composition with reasonable consistency between different organisms, can be readily extracted from soils and have wide-ranging PLFA profiles determined by gas chromatography and mass spectrometry. Indicator types associated with fungi include $18{:}2\omega6$, as a general indicator, and $16{:}1\omega5$, which is apparently specifically associated with AM fungi. Immunological techniques based on polyclonal or monoclonal antibodies, particularly via the enzyme-linked immunosorbent assay (ELISA) can be used to detect the presence of specific fungi in soil samples. Analysis of nucleic acids directly extracted from soils is revolutionizing soil ecology, including mycology. With burgeoning knowledge of the specific nucleic acid sequences associated with particular fungi, especially in relation to ribosomal DNA (rDNA; for fungi typically 28S, 18S, and 5.8S, and associated internally transcribed spacer, ITS, regions), it is possible to prescribe probes that detect fungal DNA with virtually any degree of specificity, from the broadest of scales to strain-specific. Such probes can be labeled with fluorescent, radioactive, magnetic, or enzyme tags which can enable their direct detection in soil-based systems. Application of the polymerase chain reaction (PCR) can enable the detection of extremely low concentrations of specific DNA

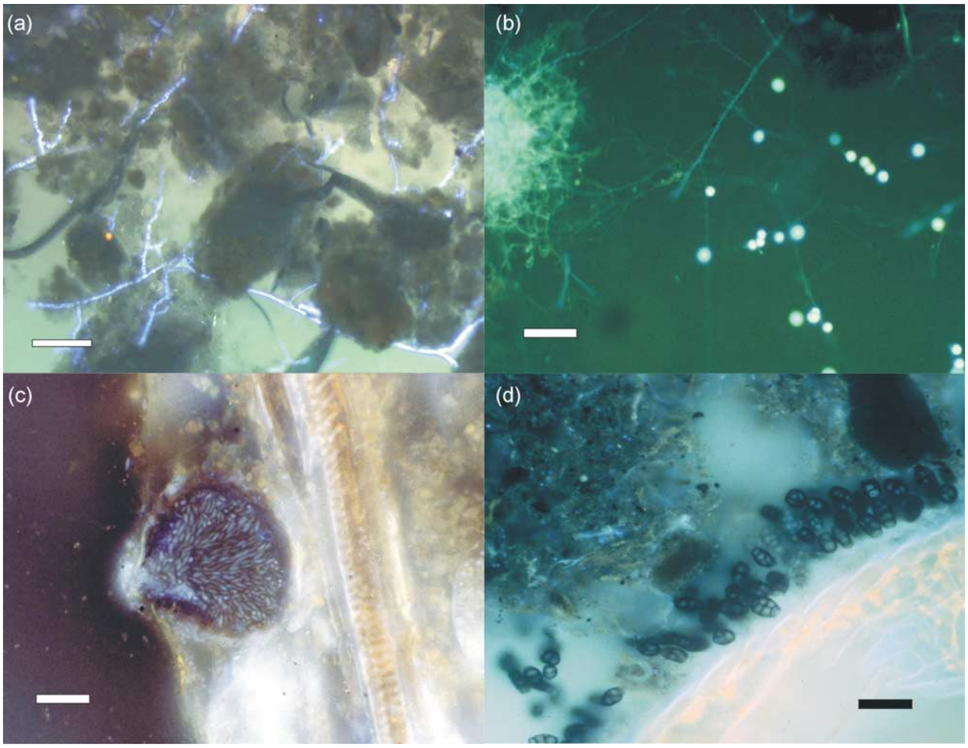


Figure 4 Fungal features visualized in thin sections of undisturbed soil. (a) Hyphae of *Rhizoctonia solani* predominating in soil pores. Scale bar 100 μm; (b) unidentified mycelium in grassland soil showing extensive branching in soil pore, and associated sporophores. Scale bar 50 μm; (c) perithecium of unidentified fungus embedded in cortex of root in thin-section of arable soil (note orientation of orifice through which spores are released into soil pore). Scale bar 50 μm; (d) accumulation of fungal spores in vicinity of decomposing root (located in lower-right quadrant of image). Scale bar 50 μm. (Reproduced with permission from Karl Ritz, David Crabb, Kirsty Harris, Scottish Crop Research Institute.)

sequences in complex environmental samples and can also be used in a broad-scale or specific sense. Many techniques exist to analyze PCR products, with cloning and sequencing providing the greatest resolution in that it literally provides the individual base-sequences of the products. It is considered that fungal DNA may be more resistant to extraction from soils than prokaryotic DNA, although there is little firm evidence for this.

Activity

Specific assessment of fungal activity in soils is difficult, since it will invariably occur in combination with the activity of the rest of the biomass. One approach is to measure bulk respiration in the presence and absence of metabolic inhibitors specific to prokaryotes such as streptomycin; assuming the inhibitors are actually specific and fully effective, the difference in respiration is attributable to fungi. More effective procedures are based on measuring the incorporation of stable or radioactive isotopes into fungal marker compounds, such as the incorporation of ^{14}C-labeled acetate into ergosterol. Since RNA is produced to a greater extent in metabolically active cells and hyphae, analysis of directly extracted RNA, in a similar manner to that used for DNA (see Molecular Methods section) offers great potential. However, RNA is difficult to extract from soils and mycological applications to date have been relatively limited.

Exploitation of Soil Fungi

Whilst soil fungi are unquestionably vital to the functioning of terrestrial ecosystems, they also play an important and increasing role in biotechnology and environmental management. Culinary exploitation and cultivation has occurred for millennia, with mushroom-growing industries present in most continents. The Périgord truffle (*Tuber melanosporum*) remains resistant to routine, large-scale cultivation and is consequently one of the most expensive epicurean foods on a weight basis, retailing at potentially thousands of dollars per kilogram. On a contrasting scale, the production of single-cell protein in large-scale fermentors from *Fusarium graminearum* is a successful and established industry. Many fungi isolated from soil provide a range of important primary and secondary metabolites which are exploited industrially. These include various organic acids (e.g., *Aspergillus*, *Rhizopus* spp.), polysaccharides (e.g., *Aureobasidium* spp.), and lipids (e.g., *Fusarium* spp.), as well as antibiotics (e.g., *Penicillium*, *Aspergillus*, *Sordaria* spp.), other pharmaceuticals (e.g., *Penicillium*, *Tolypocladium* spp.), and agrochemicals (e.g., *Strobilurus* spp.). Several thousand fungally derived antibiotics have been identified, but relatively few have found clinical use, since most are too toxic to humans. Fungi that are antagonistic to pathogens (see above) have actual and potential use as biocontrol agents against other fungi, nematodes, and weeds. Establishment of plants in horticultural or restoration scenarios can be compromised by the absence of suitable mycorrhizal hosts; inoculation with appropriate fungal partners can be effective and economically viable. Use of fungi in bioremediation of soils contaminated by industrial pollutants such as phenolics, hydrocarbons, and heavy metals has potential – claims are often made for Basidiomycete species that show strong phenol oxidase activity, such as *Phanerochaete chrysosporium* – and, in combination with plants (phytoremediation), metal-solubilizing fungi may offer effective soil cleanup mechanisms. Increasing use of soil fungi is likely given further research and understanding of the biology of these highly diverse and versatile organisms.

***See also*:** **Bacteria:** Soil; **Biodiversity**; **Fauna**; **Microbial Processes:** Environmental Factors; **Rhizosphere**

Further Reading

Cairney JWG and Burke RM (1996) Physiological heterogeneity within fungal mycelia: an important concept for a functional understanding of the ectomycorrhizal symbiosis. *New Phytologist* 134: 685–695.

Carlile MJ, Watkinson SC, and Gooday GW (2001) *The Fungi*. London, UK: Academic Press.

Davet P and Rouxel F (2000) *Detection and Isolation of Soil Fungi*. Translated by Thanikaimoni K. Enfield, NH: Science Publishers.

Dighton J (2003) *Fungi in Ecosystem Processes*. New York, NY: Marcel Dekker.

Domsch KH, Anderson T-H, and Gams W (1981) *Compendium of Soil Fungi*. London, UK: Academic Press.

Gadd GM (2001) *Fungi in Bioremediation*. Cambridge, UK: Cambridge University Press.

Garrett SD (1981) *Soil Fungi and Soil Fertility*. Oxford, UK: Pergamon Press.

Griffin DM (1972) *Ecology of Soil Fungi*. London, UK: Chapman and Hall.

Lemke PA, Esser K *et al*. (eds) (1994 *et seq*.) *The Mycota: A Comprehensive Treatise on Fungi as Experimental Systems for Basic and Applied Research*. Berlin, Germany: Springer-Verlag.

Rayner ADM (1991) The challenge of the individualistic mycelium. *Mycologia* 83: 48–71.

Smith SE and Read DJ (1997) *Mycorrhizal Symbiosis*. London, UK: Academic Press.

Wainwright M (1988) Metabolic diversity of fungi in relation to growth and nutrient cycling in soil – a review. *Transactions of the British Mycological Society* 90: 159–170.

G

GEOGRAPHICAL INFORMATION SYSTEMS

J Böhner, Universität Göttingen, Göttingen, Germany
T Selige, Technische Universität München, Munich, Germany
R Köthe, Scilands, Gesellschaft zur Bearbeitung digitaler Landschaften, Göttingen, Germany

Introduction

The term ‘geographic information system’ (GIS) has become synonymous with a wide range of computer applications, technologies, and scientific methods related to the use of geospatial information. Although the origin of GIS can be traced back to early electronic mapping tools, such as the Canada Geographic Information System (CGIS) developed in the mid-1960s, limiting the definition of GIS to a spatial visualization tool would be to disregard its current scientific and technological standards, as well as the huge expansion of GIS applications in science, politics, and economics. Driven by the rapid growth of computer technologies on the one hand and globally increasing demands on spatial information on the other, GIS has evolved from an electronic mapping facility into a fast-growing spatial science technology in less than 40 years. Due to the amalgamation of computer-assisted cartography with database technologies and growing capabilities of analyzing data across different layers in an object-oriented programming environment, current GIS provides a universal management technology for capturing, analyzing, modeling, and displaying spatial data.

Apart from its technical dimension, the term GIS also describes an emerging methodical discipline which has become an established part of the curriculum, especially in geography. However, the close connection between geography and GIS is largely based on a chance semantic coincidence and less the result of methodical commitments to computer-based spatial analyses that took place during the so-called quantitative revolution of geography in the early 1970s. The evolution of GIS was fostered by the special applications, ideas, and GIS-based solutions of scientists, federal and national agencies, governmental departments, and commercial providers, as well as the huge user community. A multitude of alternative terms for GIS bear witness to the rapid spread of GIS technology, ranging from ‘market-analysis information system’ to ‘image-based information system’. Since the early 1990s, these have included the term ‘soil information system.’

Indeed, the notion of quantitative pedology, i.e., the modeling of soil formation as a function of state factors, initially seemed perfectly suited to the new GIS technology, especially since key functional features enabled overlay and modeling operations across apparently disparate spatial data sets. Although these optimistic assessments of ‘intelligent technologies’ soon had to be revised and, particularly in soil sciences, were followed by some disillusionment, mainly on account of the sheer complexity of soil, three GIS applications to soils have now become established. These take GIS back to its original purpose as a useful instrument in: (1) data inventory and management, (2) data analysis and mapping, and (3) modeling and decision support. Before discussing these in more detail, the following will summarize major GIS applications to soils to provide an overview of elementary features, basic principles, and methods of GIS.

Elementary Features and Basic Principles of GIS Applications to Soils

Basic principles of soil-related GIS applications are best shown using a three-step scheme of GIS evolution. Based on the long-standing development of the Canadian GIS family, with its strong commitment to environmental issues, the three-stage scheme is also appropriate for showing how soils are integrated into a GIS context within the above fields of application. A structured scheme, organized along a hierarchical sequence of activities and GIS applications, is given in Figure 1. Here, the assumption is of a research process ultimately targeted at supporting management decisions.

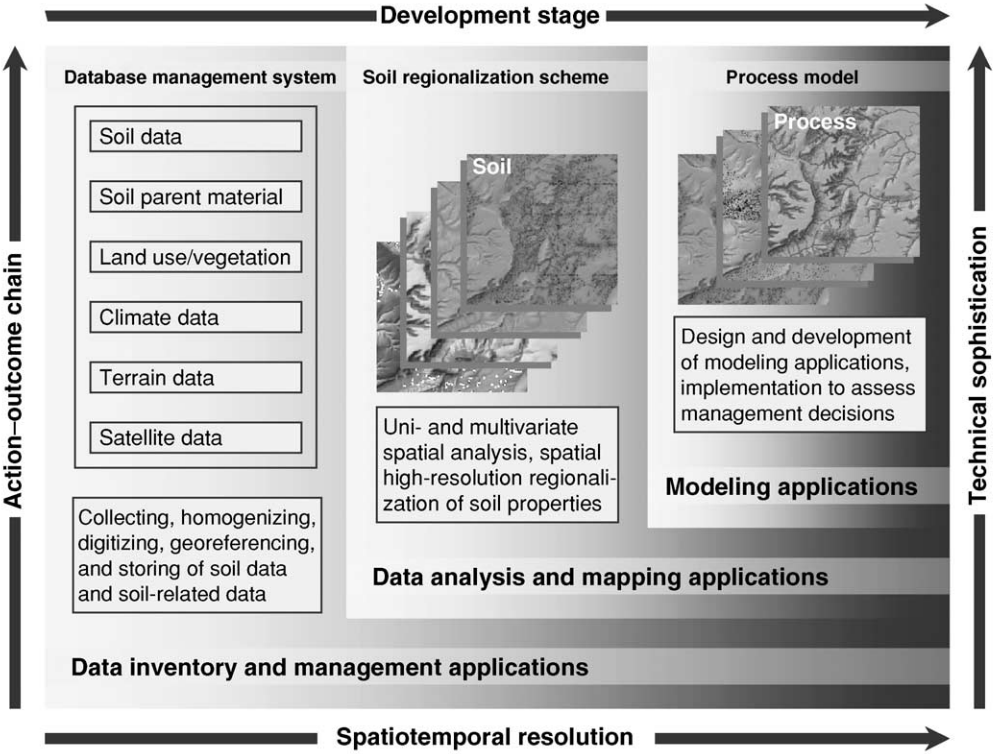


Figure 1 Conceptual model of the role of geographic information systems (GIS) and GIS applications to soils.

Taking the input data level as the essential and most crucial resource of a GIS, the initial phase of GIS evolution is characterized by data inventory and management applications. Assembling an appropriate database involves collecting, homogenizing, digitizing, georeferencing, and structured storing of data. In purely technical terms, the work consists of the design and implementation of a database management system suitable for fast data queries or searching. Although distinctions between soil and other intimately related environmental layers such as vegetation or climate are technically difficult, proper digital representation of soil and its three-dimensional variability above all requires different statistical data scales, ranging from nonquantitative nominal (e.g., soilscape) to metric entities (e.g., the organic matter content of a certain horizon). Since soils have in the past been entered into GIS as part of ambitious national or federal mapping programs that based their inventories on available soil maps, analogous soil maps have so far constituted a major data source. This leads to a distinct dominance of vector data representations, commonly consisting of discrete soil entities such as genetic soil types and local profile descriptions. Regularly spaced, continuous raster data representations of soil properties in contrast are still less common. Despite much progress in GIS-based data assembly and data management, digital maps and attribute tables therefore remain simple metaphors of the traditional soil map. The latter already represents a condensed and highly abstracted result of a knowledge-based research process, drawing on all the skills of soil scientists and cartographers at their current state of the art. Since all subsequent tasks and operations remain restricted by these academically essential but hardly operational issues, proper soil data representation demands at least some additional assembly of metadata information (e.g., site-specific descriptions of profile settings, scale issues, map conventions) if a solid scientific evaluation of all subsequent research results is to be enabled.

While inventory and management applications reflect the more technical dimension of GIS, the second stage of GIS evolution emphasizes data analysis and mapping applications. Since the term ‘spatial analysis’ includes a wide range of different methods that greatly vary in sophistication and complexity, a differentiation according to uni- and multivariate analysis to complement corresponding statistical methods is suggested. In this case, univariate spatial analyses would comprise all applications that analyze spatial (neighborhood) dependencies within one layer, e.g., autocorrelation, routing, or geostatistical analyses. Examples for soil-based applications include neighborhood analyses for delineating soil associations, or the geostatistical analyses frequently used to estimate a continuous metric soil layer (e.g., horizon depth) from random field data. The multivariate option could thus denote all kinds of analyses performed across different layers, leading to the production of a new spatial data set. This latter key functionality of a GIS is often utilized for soil regionalization purposes and will be discussed below in the Data analyses and mapping applications section. Although most current hybrid or extended GIS are capable of performing both kinds of spatial analyses using both vector and raster data, functionality and quality of analysis results still depend on which data model is used and reflect the former preferences and roots of software packages. Particularly in soil science, where spatial analyses and mapping are based on two differing views of soil, either regarding it as a spatially discrete genetic entity or a composition of spatially continuous layers, a GIS software package still needs to be chosen in accordance with the type of spatial analyses to be carried out.

In the third and most developed phase of GIS, modeling and decision-support applications are emphasized. This results from the user's desire to understand spatial patterns and processes or to undertake more complex modeling operations, required for instance to assess objectively environmental management or political decisions. Due to increasing computing capacities and advanced object-oriented programming environments, current GIS provides sufficient modeling capabilities to render this most prominent application of GIS. This in turn substantiates the use and view of GIS as a decision-support system. While soil still forms the actual research object or predictant variable in GIS-based analysis and mapping, the modeling stage integrates soil and its closely related environmental layers as predictor variables. Their aim is to model more complex climatic processes or to run scenarios of land use change in an attempt to predict the effects of management decisions.

Throughout these stages of GIS development, soils evolve from a rather static object of inventory to an essential, highly dynamic, reactive and controlling subject. In view of the spatial resolution required by certain applications, large-scale inventories satisfactory for purposes of national or federal information are replaced by a high-resolution regionalization of soil properties capable of supporting complex modeling applications. The importance of GIS also changes; it is best described as a transition process, beginning with its use as a transaction-processing system and technical tool for data management. GIS then goes on to become an array of scientific analytical methods capable of supporting the basic demands of spatial information science, with the final stage as a decision-support system, with potential uses within a range of applications.

Data Inventory and Management Applications

In view of society's growing demands for environmental services, a huge number of international, national, and federal inventory and mapping programs have been set up from the early pioneering stage of GIS in the 1960s. Initially conceived against the urgent need for increased agricultural production, more recent aims include the maintenance and qualitative enhancement of natural and rural (agro-)environments. Due to the dual role of soil as a production factor and an interacting and controlling layer within the atmosphere–biosphere–geosphere system, the assemblage of soil data was of particular interest in the context of developing ambitious agroenvironmental instruments capable of supporting sustainable development. The momentum of this development is reflected in an increasing number of soil information systems. Although they are mostly embedded in primary land or environment information systems, they distinctly show the technical and scientific specialization required for handling and providing digital soil and soil-related information.

On a global scale, the broad-resolution *Digital Soil Map of the World* (DSMW), a digital derivative of the Food and Agriculture Organization of the United Nations (FAO/UNESCO) soil map of the world, is considered the most detailed, globally consistent soil data set. The original 1 : 5 000 000 map represents a generalization of more detailed data derived from about 11 000 soil maps reviewed from various countries, which vary widely in reliability, detail, precision, scales, and methodologies. In line with the analogous source, both vector- (scale 1 : 5 000 000) and raster (spacing 5 × 5 arc min)-based derivatives include soil types, texture, profile depth, and surface slopes, as well as derived soil properties provided by interpretation programs and related data files on agronomic and environmental parameters. Apart from the maps of classification units contained in the *World Soil Reference Base* (WSRB) and tables on special soil analyses for every country of the world, an additional soil database was created for global environmental studies, including data on soil-moisture storage capacity, soil-drainage classes, and effective soil depth. A corresponding data set of water-holding capacities in a $1^\circ \times 1^\circ$ grid (latitude, longitude), drawn from original sources to meet the requirements of general circulation model (GCM) experiments, can be quoted as another example of using soil data in the context of global environmental studies.

Despite its broad spatial resolution and distinctly limited information content, features that are unlikely to be improved by the current SOTER revision (soil and terrain database of the world), the compilation of the global soil map and its digital derivative did contribute much to academic debates on the global harmonization of soil classification concepts and may also have influenced the development of many continental, national, and federal soil information systems. Prominent examples of well-established projects, such as the National Soil Information System in the USA (NASIS), the Canadian Soil Information System (CanSIS), the Australian Soil Resource Information System (ASRIS), or the European Soil Information System (EUSIS), form a suitable basis for a steady, coordinated, and harmonized inventory of soil and soil-related data in accordance with essential environmental tasks. Indeed, precise manuals of procedure setting out requirements, inherent concepts, and methods essential for a rationalized inventory of georeferenced, digital soil databases have given rise to an increasing availability of plausible and consistent soil data accessible for many purposes.

Using the ambitious EUSIS project as an example, Figure 2 shows the construction of a soil information system database. The relational structure consists of topological, semantic, and geometric data sets, organized along a hierarchical chain of discrete soil entities (objects). The topological tables contain information on the distribution of soil bodies within soilscapes and soil horizons within soil bodies (organization tables), as well as information on the nature of limits, separating different soil-scapes within a soil region (limits table). The properties of soils, mostly attached to soil bodies and horizons, are given in the semantic attribute data sets, whereas positions and shapes are captured by the geometric-attribute data sets. To address the needs of soil information users at different scales, EUSIS is organized along a scale-dependent structure. Spatial soil information is assembled and provided at global 1 : 5 000 000 scales compatible with the SOTER database of the FAO, at intermediate 1 : 1 000 000 scales covered by the *Soil Geographical Database of Europe* (SGDBE) and at 1 : 250 000 scales. Larger scales (1 : 50 000, 1 : 5 000) are covered by regional to local authorities. To ensure homogenous data contribution and assembly, EUSIS partly links up with existing national and federal soil information systems (e.g., Lower Saxony Soil Information System, NIBIS, Germany). Special working groups in the European Soil Bureau (1 : 1 000 000, European Soil Database working group; 1 : 250 000, working group) coordinate data contribution and management in accordance with the relevant scales.

Despite clear differences between individual soil information systems, particularly in terms of subdividing soil categories and the match between soil taxonomy and soil map units, most systems reflect the EUSIS structure and are conceptualized along traditional map scale conventions. Since the genetic soil category or soil type, although somewhat modified, is still employed as the major response unit, the common practice of direct, rule-based delineation of pedotransfer functions from these discrete entities yields no more than a coarse estimate, with clear limitations in terms of spatial distribution, resolution,

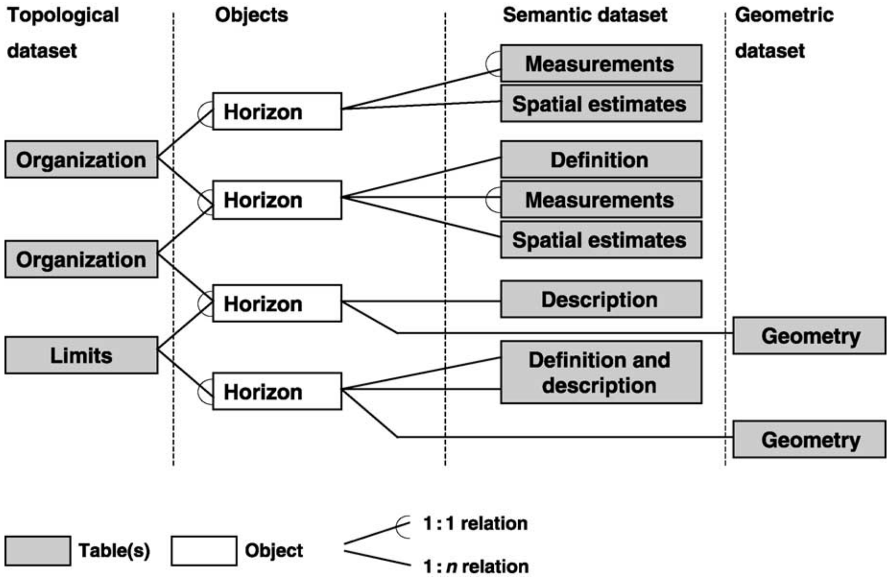


Figure 2 General structure of the European Soil Information System (EUSIS) database.

and reliability. Though soil information systems technology stresses its developments by catchwords such as 'multiuser access' or 'just-in-time performance,' its continued focus on discrete mapping units may well satisfy classification and pedogenetic criteria but does not necessarily meet the needs of prospective uses. Despite the global increase in reliable soil data, the current global soil inventory is not in a position to meet the increasing demands posed by soil protection, sustainable soil management, or other major political issues. These ambitious aims demand a spatially flexible database of different soil properties, which may be obtained through further GIS analyses.

Data Analyses and Mapping Applications

GIS-based data analyses and mapping operations have been widely developed for the purpose of yielding high spatial-resolution estimates of different soil properties. Apart from univariate interpolation techniques performed on random field data, for example, multivariate regionalization approaches in quantitative soil pattern analyses are the most frequently used, describing the spatial variability of soils in terms of a set of determining predictant variables and thus providing insight into the processes and factors involved. The term 'regionalization' is preferred in this context, since it comprises the somewhat misleading terms of 'upscaling,' 'downscaling,' 'interpolation,' and 'extrapolation' and also covers the major contemporary approaches of soil-mapping support, roughly to be divided into a predominantly vector-based approach and a continuous raster-based approach.

The knowledge-based discrete approach (or concept map approach) is still the most frequently used instrument of support in soil mapping. In accordance with Jenny's factorial theory of pedogenesis, a set of state factors such as, e.g., parent material, terrain, climate, or vegetation is statistically compared with the spatial soil pattern by utilizing overlay and intersection algorithms as key functional operations of GIS. The smallest common geometries of state factors are identified as a surrogate for the site-specific process structure, to enable a rule-based assignment of soil parameters and infer estimations of spatial soil patterns with a certain probability. Since state factors within a landscape system are intimately related, a landscape component can thus be assigned to typical patterns of landform elements, soil types, or vegetation sequences at each level of the spatial hierarchy. Adequate capture of covariance structures allows data gaps to be filled (e.g., in the ASRIS project performed on the basis of adjacent, properly mapped soil distributions), or infers spatial soil pattern at finer resolutions commensurate with terrain or vegetation information. By integrating sophisticated complex analytical methods such as fuzzy logic or the use of neuronal nets, characterized by their reduced reliance on expert knowledge in computing concept maps, descriptive rule-based models and expert systems can yield many supporting features of an improved and extended spatial definition of soil patterns. However, the discrete approach has distinct limits, particularly resulting from the finite number of predictant entities which may neither represent the spatial soil diversity in the area concerned nor the process scopes relevant to soil. Also, the classification of continuous, process-controlling geofactors (e.g., climate, relief) represents a critical task, since it involves an a priori assumption that the limits inferred are actually relevant for the distribution of soil characteristics and its spatial definition.

In view of these deficiencies, the continuous soil regionalization approach is sometimes preferred, particularly where metric soil properties are required for further use, e.g., in process models. Based on the definition of soil as a composition of continuous layers, a set of continuous terrain parameters has been used as predictor variables in the context of using statistical and geostatistical procedures for inferring spatial soil estimates from metric point data. Results indicate significant correlations between soil and terrain variables, particularly when complex, hydrologically based terrain indices, such as the Terrain Wetness Index (TWI; see Figure 3) are analyzed. Other results confirm that much of the variation apparent in soil is a response to near-ground runoff and associated translocation processes. To ensure consistent applicability of a continuous regionalization scheme to other climatic regions, recent applications have extended the continuous approach by integrating spatial high-resolution climate and terrain variables to complex process parameters. Figure 3 exemplifies the Mass Balance Index (MBI) for the parameterization of slope transfer processes, integrating the universal soil-loss equation (USLE) *R*-factor and different catchment area parameters (catchment area, catchment slope). Since terrain parameters based on nonlinear discharge models tend toward strongly artificial variability, particularly in hydrologic homogenous areas (e.g., large valley grounds), the SAGA (system for an automated geoscientific analysis) Terrain Wetness Index (STWI) applies a slope-dependent, iterative modification scheme to the catchment area size (Figure 3) which keeps it suitable for the regionalization of hydromorphic soil properties. Different statistical and geostatistical regionalization methods performed on a set of terrain and process parameters were evaluated with regard to their routine application in the context of soil mapping. Universal kriging

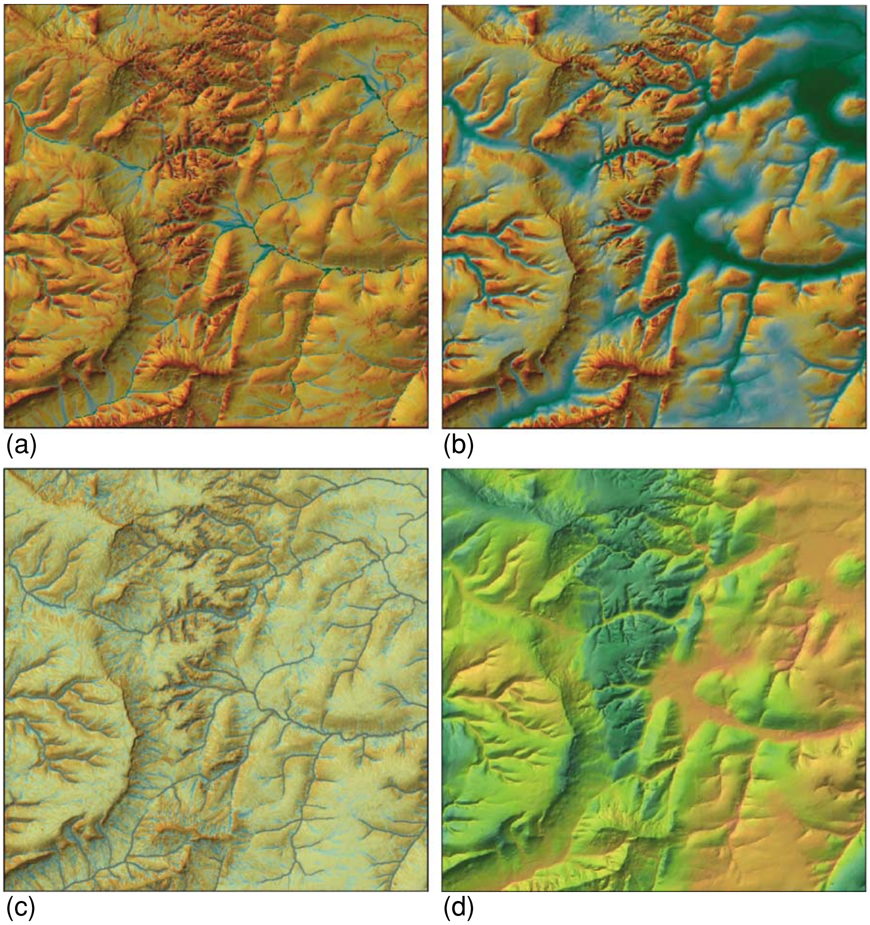


Figure 3 (a) Terrain wetness index. (According to Moore ID, Gessler PE, and Peterson GA (1993) Soil attribute prediction using terrain analysis. *Soil Science Society of America Journal* 57: 443–452); (b) SAGA (system for an automated geoscientific analysis) terrain wetness index; (c) mass balance index; and (d) silt content of the soil surface layer (red, high silt content; green, low silt content). (Adapted from Böhner J, Köthe R, Conrad O *et al.* (2003) Soil regionalization by means of terrain analysis and process parameterization. In: Micheli E, Nachtergaele F, and Montanarella L (eds) *Contributions to the International Symposium Soil Classification 2001,* Ispra, Italy. EUR 20398 EN. European Soil Bureau, Joint Research Centre.)

was assessed as a valuable tool for soil mapping and further modeling support. Results are exemplified for the regionalization of the topsoil silt content obtained by universal kriging (Figure 3). The strong emphasis on process throughout the regionalization scheme reflects the basic assumption that no matter which strategy is used for soil regionalization, a reliable spatial prognosis of soil properties can only be achieved if the underlying spatial data directly represent soil-related processes rather than being substituted by discrete factor combinations. Distinct limits in applicability, often the main focus of criticism, are due to the comparably high demands on soil measurements or profile data. In view of the growing availability of very-high-resolution remote sensing and particularly radar data, capable, for example of picturing topsoil moisture distribution and thus topsoil subscale (random) variability, this difficulty is likely to be overcome in the medium term. Nevertheless, a proper point database remains an essential basis for continuous soil regionalization.

Modeling and Decision-Support applications

Since the above soil regionalization approaches empirically detect predictable set structures within spatial soil variations and are thus occasionally termed 'models' themselves, this term is really reserved for the type of process modeling applications described below. The integration of soil as a reacting and controlling body into process models expands the capabilities of quantitative, GIS-based simulation, with ensuing benefits to a range of applications, particularly in the context of management decision support. Apart from hydrologic models, with their specific demands on horizontal and vertical soil data distribution, a wide range of applications exist in the context of soil erosion risk assessments. They can roughly be divided into empirical and physical approaches, with both requiring pedophysical values or at least a parameterization of certain soil properties.

Possibly the most prominent approach for estimating water-induced soil erosion, the USLE, is now predominantly carried out on a GIS basis on account of its comparatively simple structure. The USLE consists of an empirical equation which calculates the mean annual soil-erosion rates on agricultural slopes as a function of six variables (R, precipitation factor; L, slope length factor; S, slope factor; C, vegetation cover and tillage factor; K, soil-erodibility factor; P, erosion-protection factor). The parameterization of the topsoil erodibility in the K-factor considers soil particle size distributions (sand and silt percentages), organic matter content, and classes of permeability of the soil surface layer, e.g., entailed or delineated from soil taxonomy profiles. Comparable demands on the parameterization of topsoil properties are also valid for the empirically based wind-erosion equation (WEQ), predicting the potential mean annual soil loss at the field scale by determining the influence of several primary variables characterizing soil erodibility, wind erosivity, and the actual land-use situation.

Due to the limited transferability of empirically based erosion risk assessments, more sophisticated, process-based approaches have recently been developed capable of simulating water and wind erosion in a physically consistent manner. Their major advantage of wide applicability is countered by the disadvantage of extended data requirements on the spatiotemporal resolution and the quality of input parameters. This is particularly valid for soil data, which

Figure 4 Spatial distribution of the erosion–accumulation balance for the event of 13 and 14 March 1994 at the Barnham test site, UK. Positive values (green) indicate accumulation; negative values (yellow and red) indicate erosion (dark green, forests; gray, settlements and Honington airport). Average wind direction of the storm was west (240–300° direction angle). Model domain covers 5 × 5 km. Reproduced from Bohner *et al.* (2003) The Weels model: methods, results and limitations. *Catena* 52: 289–308, with permission from Elsevier.

have to be available either as physical soil attributes or as delineated parameterizations. With respect to its prospective use in case studies, management decision support, and particularly in context with precision farming applications, major requirements concern the spatial resolution and thus demand a suitable procedure for the regionalization of metric topsoil properties.

Emphasizing this, Figures 4 and 5 exemplify the results of physically based wind-erosion simulations using the wind erosion on European light soils (WEELS) model, performed on spatial high-resolution estimations of soil particle distribution and surface-layer organic matter content. Apart from the soil-erodibility factor, soil attributes are essential inputs for the temporal high-resolution simulation of topsoil moisture contents, since they essentially determine the process of wind erosion. The WEELS model also considers various climatic parameters, annual changes in land use, crop cycles, and crop phenology, as well as the influence of the terrain. Due to its high temporal resolution (hourly) the WEELS model allows the simulation of wind-erosion events (Figure 4 shows an event at the Barnham test site, UK) as well as long-term estimates of erosion and accumulation rates. The spatial distribution of erosion and accumulation rates for the Barnham test site (Figure 5) was derived from a 29-year model run (1970–98). The option of long-term simulations enables the assessment of different land-management scenarios and their susceptibility to wind-erosion processes. The result of land-use change scenarios for a test area in Grönheim in the Cloppenburg Geest area (the western part of Lower Saxony) has shown the possibility of minimizing erosion risks through sustainable land-use strategies. This can underline the capabilities of GIS-based modeling and decision support in the context of environmental policy and management.

Figure 5 Modeled spatial distribution of the mean annual erosion–accumulation balance at the Barnham test site, UK (1970–98). Due to changing wind directions and annually varying vegetation cover, the orientation of the soil erosion rate gradient may differ among the individual fields (dark green, forests; gray, settlements and Honington airport). Model domain covers 5 × 5 km. Reproduced from Bohner *et al.* (2003) The Weels model: methods, results and limitations. *Catena* 52: 289–308, with permission from Elsevier.

Summary

The evolution of GIS in terms of technical maturity and functionality offers many possibilities for application to soils, ranging from data assembly and inventory support to complex modeling. The proliferation of GIS within the scientific community has fostered its frequent use in spatial data analyses, particularly in the context of soil-mapping support and soil regionalization. Due to their increasing ability to analyze data across different layers, GIS constitute suitable instruments for developing soil-related process models with wide potential applications for instance in decision-support systems. However, the clear technical advantages cannot detract from the fact that the most crucial resource of a GIS, the database itself, still tends to be a digital derivative of a traditional soil map. It therefore comprises individual scientific field experience as well as mapping conventions, neither of which can be easily reproduced in a digital format. The persistance of the genetically well-founded, discrete soil entity as the basic information unit and particularly the direct delineation of (coarse estimated) pedotransfer functions is hardly capable of meeting the increasing demands for soil information in terms of spatial resolution and reliability, and thus to a certain extent even violates the capabilities of current GIS. A major alternative, the spatial high-resolution regionalization of available soil profile data – an alternative which would deliver more-reliable soil estimates despite its being transcribed from an older classification system – above all requires an appropriate soil profile database. Though soil profile data are often compiled as part of developing national or international soil information systems, a systematic compilation and homogenization of available profile data on a national or international level is currently unavailable. This is therefore suggested as one of the most important future tasks in the further development of national and international soil information systems. Clearly this is no easy feat, but a rewarding task in view of the expected increase in knowledge on soil-formation processes. Rather than GIS users, it will, however, require GIS-using soil scientists.

See also: **Environmental Monitoring**; **Erosion:** Water-Induced; Wind-Induced; **Remote Sensing:** Organic Matter; **Spatial Variation, Soil Properties**; **Statistics in Soil Science**

Further Reading

Böhner J, Köthe R, Conrad O *et al.* (2003) Soil regionalisation by means of terrain analysis and process parameterisation. In: Micheli E, Nachtergaele F, and Montanarella L (eds) *Soil Classification 2001* EUR20398 EN. Luxembourg: Office for Official Publications of the EU.

Böhner J, Schäfer W, Conrad O, Gross J, and Ringeler A (2003) The WEELS model: methods, results and limitations. *Catena* 52: 289–308.

Bullock P, Jones RJA, and Montanarella L (eds) (1999) *Soil Resources of Europe*. EUR18991EN. Luxembourg: Office for Official Publications of the EU.

Crain IK and MacDonald CL (1984) From land inventory to land management. *Cartographica* 21: 40–46.

FAO–UN (1970–78). *Soil Map of the World*, scale 1 : 5 000 000, vols I–X. Paris, France: UN Educational, Scientific, and Cultural Organization.

Finke P, Hartwich R, Dudal R *et al.* (1998). *Georeferenced Soil Database for Europe – Manual of Procedures*, version 1.0. EUR18032EN. Luxembourg: Office for Official Publications of the EU.

Gessler PE, Moore ID, McKenzie NH, and Ryan PJ (1995) Soil landscape modelling and spatial prediction of soil attributes. *International Journal of Geographical Information Systems* 9: 421–432.

Longly PA, Goodchild MF, Maguire DJ, and Rhind DW (1999) *Geographical Information Systems*, vol. 1, *Principles and Technical Issues*. New York: John Wiley.

Maguire DJ, Goodchild MF, and Rhind DW (1991) *Geographical Information Systems – Principles and Applications*, vol. 1, *Principles*, pp. 9–20. Harlow, Essex: Longman.

Moore ID, Gessler PE, Nielsen GA, and Peterson GA (1993) Soil attribute prediction using terrain analysis. *Soil Science Society of America Journal* 57: 443–452.

Morgan RPC, Quinton JN, Smith RE *et al.* (1998) The European soil erosion model (EUROSEM): a process-based approach for predicting sediment transport from fields and small catchments. In: *Earth Surface Processes and Landforms*, vol. 23, pp. 527–544. New York: Kluwer.

Schmidt J, von Werner M, and Michael A (1996) *Erosion 2D/3D – ein Computermodell zur Simulation der Bodenerosion durch Wasser.* Leipzig: Sächsisches Landesamt für Landwirtschaft.

Sinowski W (1995) *Die dreidimensionale Variabilität von Bodeneigenschaften.* Forschungsverbund Agrarökosysteme München (FAM), vol. 7. Aachen: Shaker-Verlag.

Wischmeier WH and Smith DD (1978) *Predicting Rainfall Erosion Losses – A Guide to Conservation Planning.* Agricultural Handbook No. 537. Madison, WI: USDA.

Woodruff NP and Siddoway FH (1965) A wind erosion equation. *Proceedings of the Soil Science Society of America* 29: 602–608.

GERMINATION AND SEEDLING ESTABLISHMENT

A Hadas, The Volcani Center, Bet Dagan, Israel

Introduction

Seed germination and seedling establishment are critical phases in the plant growth cycle, since they influence and determine species survival in natural habitats and the onset and yields of agricultural crops. A dormant seed combines an embryonic plant and stored materials to be used during germination and seedling establishment until the onset of photosynthesis by the established seedling. 'Germination' is a general term describing the sequences of complex processes involved in initiating an array of metabolic activities that lead eventually to the renewal of growth of the dormant seed embryo and, ultimately, to seedling establishment. The various processes known to occur during these two phases take place in parallel or in serial sequences. Their times of initiation and of transition from one sequence to another are triggered by events endogenous to the seeds (seed-development processes of the parent plants) and depend greatly on the prevailing environmental conditions (aeration, temperature, available water and nutrients, soil mechanical impedance) and seed-development conditions.

Seed Germination and Environmental Conditions

Proper seed germination and stand establishment depend strongly on the environmental conditions (moisture, thermal and aeration regimes in the soil, light). Where favorable environmental conditions prevail, other factors may decide the success or failure of seed germination and stand establishment (seed-development processes on the parent plants, seed dispersion, and depth of seed burial). Seeds are self-contained units, owing to the storage materials contained in the seeds. Therefore, the environmental requirements for germination and seedling establishment are considered to be fewer and simpler than those required by full-sized plants.

Water Requirements

Water uptake by seeds is a prerequisite for proper germination and seedling establishment. The amount of water required by a seed for germination itself is considered to be very small, but the rate of uptake and the total amounts taken up depend greatly on the seed and soil properties with respect to water – the water potential differences between the seed and the soil – and are controlled by the water conductivity of the soil and the soil–seed contact zone. Whether or not the amount of water taken up will suffice for germination and seedling development depends on the water energy status of the seed, and on the soil-water potential. There is a certain minimal level of seed hydration, termed the 'critical hydration level,' below which the seed will not germinate; instead it may enter secondary dormancy. Seeds that enter secondary dormancy become more susceptible to pests and diseases. These seeds may germinate later in the season or during the coming seasons. Thus the uniform germination, establishment, and stand of crops are impaired. The viability of the seeds that reenter secondary dormancy diminishes with time. Species and cultivars differ markedly in conditions they require for germination, and these differences are attributed to the differences among the water regimes and other soil physical conditions of the soil to which the plants were adapted or which they encountered during germination.

Temperature Requirements

Temperature affects the soil properties, with respect to water, that influence seeds, air and water regimes, and the biological activity of seeds themselves. Soil temperature varies greatly, because of both diurnal and seasonal thermal processes, and is dependent on the constituents of the soil (water, solids, organic materials, and air), its texture, structure, color, and layering. For germination to occur, the temperature of the seed environment should fall within a favorable, species-specific range. Within that interval lies the optimal temperature at which optimal germination is observed. The temperature limits (minimum and maximum temperatures) below or above which no germination will occur are 3–5° and 30–43° for wheat cultivars; and 8–12° and 40–44° for corn cultivars. The corresponding optimal temperatures are 15–30° and 26–32° for wheat and corn cultivars, respectfully. Planting during spring depends on the soil temperature and, as an example, corn will be planted only when the soil temperature at the planting depth, during most of the diurnal cycle, is slightly above its minimal temperature for germination. Favorable temperature ranges, specific germination-enhancing periodicities of diurnal or seasonal temperature variation, induction of secondary dormancy, and the combined effects of water stress and temperature vary among species. Germination is greatly affected by the interactions among temperature, soil-water potential, and water flow in the soil, and by variations in the Q_{10} factors of effective rates of seed biological activity. Each of the processes that occur in plants has its own rate response to temperature, commonly designated as a coefficient, Q_{10}, which is defined as the number of times that process rate increases with a 10° rise in temperature. The plant response to interactions between temperature and processes involved in water uptake, respiration, or cell division and growth are extremely complicated and unpredictable. Adverse effects of soil-water stress on germinating seeds intensify as temperature rises and may persist beyond the germination stages, into the seedling growth and emergence stages. Under field conditions, soil temperature is characterized by a diurnal temperature fluctuation, with amplitudes as great as 10° or more. If seeds respond to a single process, the observed rate will differ from that measured under constant temperature conditions.

Aeration Requirements

The total air content in the soil and the rate of gaseous exchange greatly affect the soil biological activity and the availability of oxygen to germinating seeds and developing seedlings. However, the effects of the interactions between the total air content, its constituents and composition in the soil, and the rate of gaseous exchange are complex and difficult to define, unless a definite knowledge of the interrelationships among complex diffusion processes (in air-filled pores and across water films) that control the oxygen supply and the dissipation of respiratory and decomposition by-products (CO_2, N_2, NO_2, H_2S, ethylene, methane) is attained. Low oxygen availability reduces or even prevents germination in most species. Oxygen supply to support metabolic activity in germinating seeds becomes decisive at a very early stage of germination, namely at the beginning of respiration and utilization of storage materials, and especially when growth is initiated during seedling development. Oxygen requirements increase with soil temperature, and under light and/or water stress. Very often a conflict develops between oxygen supply and water supply to germinating seeds, because of the very low solubility and diffusivity of oxygen in water. Thus, oxygen supply is greatly impaired as the thickness of the water films around the seeds and the hydrated seed coat increases, especially with seeds that have a swollen, hydrated mucilaginous cover. Low CO_2 concentrations may stimulate germination, either solely or at times in combination with ethylene. Soil seals or crusts and compacted soil layers may have deleterious effects on gas exchange and, in turn on seed germination, and seedling development and establishment.

Light Requirements

Sensitivity to light is observed in germinating seeds of various species, which fall into four classes: those that are insensitive to light, like most agricultural crops (e.g., anemone, oregano); those that require a short exposure to light during their germination (e.g., nasturtium, compass lettuce); those that will germinate in total darkness (tulip, gladiolus, Nigella); and those that will germinate under full exposure to light. The sensitivity of light-sensitive seeds increases as they take up more water. Germination of light-sensitive seeds is impaired when they are shaded by a dense plant canopy or when light is prevented from reaching the seeds because of burial under the soil surface or in cracks in the soil.

Soil Mechanical Constraints

Soil particles of various sizes and origins form a matrix that exhibits a degree of resistance under mechanical stress, described as ‘soil mechanical strength.’ Soil strength depends on the soil constituents, density, moisture content, and structure; it increases with increasing soil bulk density, and decreases with soil-water and organic matter contents. Silty soils with low organic matter content tend to deform plastically and to compress easily, and to form surface seals under the impact and slaking action of raindrops or under instantaneous flooding by rain or irrigation. These thin, rather dense seals may restrict gaseous exchange and water infiltration into the soil around the seeds, mechanically impede or obstruct germination and seedling emergence, or there can be a combination of these effects. Wheat seedlings are affected by the interaction between crust strength, water content, and rainfall. Adverse effects on seed germination and seedling establishment, similar to those of soil seals, are caused by mechanical compaction of the soil surface.

Water Uptake by Seeds and Seedlings

Water uptake by a germinating seed is an essential step toward rehydration of its tissues. The initiation of the array of metabolic processes in the seed and the minute amounts of water required for germination depend on the seed genome and the constituents of its individual parts. The various organs (embryo, cotyledons) and tissues differ in their internal physical structure, biochemical properties, and composition; therefore, they differ in their water retention, distribution, and swelling. In most seeds the seed embryos are minute (the embryo weight varies between 0.005

and 0.02 of the total dry seed weight). Storage tissues constitute the major volume of the seeds, and the observed amount of water taken up by those seeds is that which is taken into these tissues under the water regime that exists during the germination period of these seeds.

Water uptake by dry seeds during germination is characterized by three phases: the initial phase – the imbibition phase – is characterized by a saturation kinetics pattern; the second phase – the transition phase – is characterized by a low-to-negligible water uptake rate; and the third phase – the growth phase – is characterized by a rapid, exponential increase in the water uptake rate, accompanied by the emergence of the radicle. The first two phases are observed in dead, inert, and viable seeds alike, whereas the growth phase is unique to viable, germinating seeds. Water uptake rates during these phases are controlled by one or more of the following factors: (1) the seed properties with respect to water (the seed's water content–water potential characteristics; the seed's diffusivities to water, which range between 1.7×10^2 to $1.5 \times 10^{-6}\,\mathrm{m^2\,day^{-1}}$); (2) the soil-water properties (soil-water characteristics, diffusivities to water of the soil around the seed, which range from 4.0×10^4 to $5.0 \times 10^7\,\mathrm{m^2\,day^{-1}}$); and (3) the seed–soil interface properties with respect to water (seed–soil contact area, impedance to water flow across the contact zone). Once the third phase starts, the radicle emerges, the roots elongate, and the seedling starts to take up water with its radicle and developing roots, while the hypocotyl or mesocotyl emerges and elongates toward the soil surface. Water uptake by the radicle or the roots is controlled by one or more of the following factors: (1) the soil-water properties; (2) the root anatomy (root and/or radicle structure and tissue arrangement); (3) rate of elongation, radicle or root–soil contact, and hydraulic impedance to water flow across the seed–root contact zone.

The Imbibition Phase

The imbibition phase starts with the entry of water into the seed. This water is distributed in crevices, cracks, and flaws in the seed cover, and is absorbed by the seed tissues. Water uptake rate measurements taken during this phase have shown these rates to be: (1) temperature-dependent; and (2) accompanied by increases in respiration rate and in light sensitivity in some seed species. These observations suggest that water uptake during imbibition is not a 'passive' process, as it is usually taken to be, but becomes an active one at an early stage of this phase. The end of this phase is marked by an asymptotic approach to a final water gain, or hydration level, which depends on ambient soil-water potential, soil conductivity to water, seed–soil contact, and seed composition.

The Transition Phase

During the transition phase, known also as the 'pause phase,' the seed-water content, respiration rate, and apparent morphology remain almost unchanged. Nevertheless, a variety of metabolic processes are activated, and differences in the activity levels of these processes and in their order of occurrence have been observed among seeds of different species and among seeds that have reached different hydration levels. Therefore, any adverse environmental conditions that lead to redrying of the seeds, e.g., by subjecting them to water stress, or reduce the seed–soil contact area, and thus influence their hydration levels, may impair, retard, or even inhibit germination. If no damage had resulted, no secondary dormancy been induced, and no inhibitory processes been blocked, the germination of these seeds upon rewetting would be enhanced because of the high concentration of unused metabolites accumulated prior to drying. The duration of the transition phase influences the initiation time and the extent of radicle growth. Seeds have been observed to reach the transition phase and to remain in it for long periods that may extends for days, weeks, or more before germination. Toward the end of the transition phase, the embryonic cells of the radicle start to divide.

The Growth Phase

The growth phase starts with an increase in the respiratory rate, initiation of cell division, and extension of the embryonic radicle and ends with radicle protrusion. The common definition of germination states that a seed has germination once its radicle has broken out of the seed coat and started to elongate. From that stage, with germination completed, the seedling development and growth have commenced, and water is taken up solely by the elongating roots.

Differentiation Between the Seed Germination Phases

The definition of and the differentiation among the phases are based on an arbitrary partitioning of the continuous sequential order of processes in the germinating seed and the developing seedling. In reality, all the phases are interdependent, and the interrelationships among them suggest that each phase depends greatly on the preceding phases. The end of the third phase (marked by radicle protrusion), termed 'the growth phase,' cannot logically be separated from the subsequent seedling growth toward its protrusion through the soil surface and its establishment. The end of the transition phase depends on whether the amount of water finally taken up at the end of imbibition has reached the required minimal level. A generally accepted concept is that to germinate a seed must reach a minimal water content – the 'critical hydration level.' That 'hydration level' does not reflect the water distribution among the seed components, nor is it an absolute value, since it depends on the water uptake rate, variations in soil-water potential, conductivity to water and temperature, and seed adaptability to variations in these environmental factors. The concept of critical hydration level fails in cases of partial seed wetting, especially when the wetted seed volume includes only the embryo and the adjacent storage tissues, and an affirmative concept is the 'critical water potential,' taken as the external water potential at or below which seeds cannot reach their critical hydration level. These concepts are applied respectively, when water uptake is evaluated according to its final amount of water taken up or in terms of seed–soil or seed–substrate water potential equilibrium states.

Physical Principles of Water Uptake by Seeds and Seedlings

Water movement within the soil toward the seed or radicle–soil contact interfaces, across them and into the seed or into the root, can be described by the general flow equation combining the Darcy flow equation and the continuity equation based on the principle of conservation of matter:

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial S}\left[k(\psi)\frac{\partial \psi}{\partial S}\right] = \frac{\partial}{\partial S}\left[D(\psi)\frac{\partial \theta}{\partial S}\right] \qquad [1]$$

For water movement in the soil and to the seed, it is simpler to use the right and the left sides of this equation, and the proper initial and boundary conditions. The initial conditions describe the initial state of the system under consideration, whereas the boundary conditions express the conditions that are imposed on or those that will exist once the processes to be analyzed are started. For a planted seed that starts to imbibe water from the soil around it, the initial conditions are:

$$\theta = \theta_{\text{initial}}$$
$$t < 0$$
$$0 < r \leq a$$

The boundary conditions that will exist once imbibition starts are:

$$D_{\text{seed}}(\partial \theta_{\text{seed}}/\partial r) = (LD_{\text{seed}}/a)[\theta_{\text{final}} - \theta_{\text{t}}]$$
$$t \geq 0$$
$$r = a$$

where θ, θ_{initial}, θ_{final}, θ_{t}, θ_{seed}, are the volumetric water contents of the soil, the initial and final values (cubic centimeters per cubic centimeter), at time t, respectively; θ_{seed} is the volumetric water content of the seed (cubic centimeters per cubic centimeter); ψ is the soil-water potential (kilopascals); $k(\psi)$ is the soil capillary conductivity (centimeters per day); $D(\theta)$, D_{seed} are the diffusivity coefficients of water in the soil and the seed, respectively (centimeters squared per day); r, a are the radial distance from the center of the seed and the seed radius, respectively (centimeters); and L is the seed–soil interface impedance to water flow (per day) (Figure 1). The rate of flow depends on: the water potential gradients (the driving force that induces water movement) in the soil, across the seed–soil contact zone, and in the seed tissues; and on the respective conductivities to water of the soil, the seed and radicle–soil contact zone and the seed or radicle tissues. In order to simplify the solutions to the above equation and the initial and boundary conditions, the number of variables has to be reduced. The choice to use the right side of the equation given above (where use of diffusivity to water in the soil is made) simplifies the calculations by reducing the number of variables and makes the measurements required to determine the flow parameters for use in the equation practicable.

Under normal conditions water will tend to move from the higher (in the soils) to the lower (in the seed

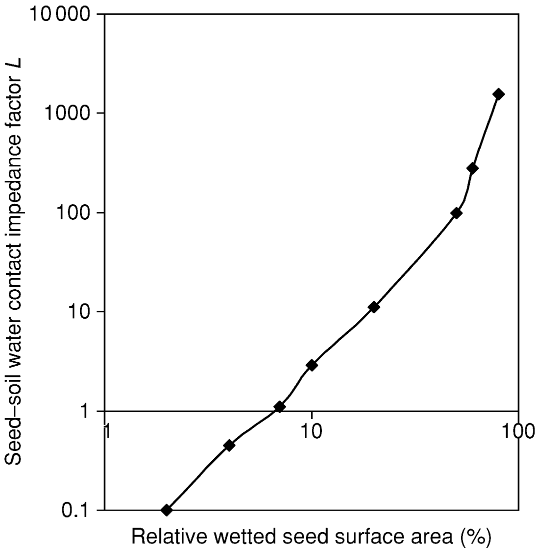


Figure 1 Seed–soil water contact impedance factor L as a function of the relative wetted seed surface area.

or radicle) water potential. The energy status of soil water termed 'soil-water potential' depends on the soil texture, structure, water content, soil solution composition, and solutes concentration. The matric potential is the soil-water potential component that arises from the interaction of soil water with the matrix of the solid particles in which it is contained. The osmotic potential is the soil-water potential component that arises from the interaction of soil-water solutes' composition and concentrations with the active membranes of the seed cells. Soil-water potential per unit of soil is given as work per unit of soil volume, or soil weight. At the beginning of the imbibition phase, these gradients are very large and the uptake rate is high, but, as the seed gains water, the gradient diminishes and the rate of water uptake decreases. As water moves into the seed, the water content in the soil next to the seed surface decreases; the matric and the osmotic components of the soil water decrease also (due to reduction of water content and an increase in solute concentration caused by ion exclusion). Consequently, the conductivity and diffusivity to water, through the soil and across the seed–soil contact, decrease, which tends to make the impedance to water flow increase. But, on the other hand, as the seed water content increases, the seed swells and the seed–soil contact improves, which tends to cause a decrease in the flow impedance. The combined effect leads to the net water flow rate into the seed.

Specific Effects of Matric and Osmotic Soil-Water Potential Components

Both the matric and the osmotic components of the soil-water potential are directly involved in soil-water movement to germinating seeds. Seeds immersed in solutions and in saturated soils respond equally to equal changes in these two components, provided their membranes are intact and fully active. But in soils, small reductions in soil-water matric potential have been observed to affect germination, and seedling development and establishment to a greater extent than equal or even greater reductions in the soil-water osmotic potential (Figure 2). This difference in response is because the changes in the seed–soil contact area and soil-water content that accompany small changes in the soil-water matric potential lead to proportionally greater changes in the soil conductivity and diffusivity to water.

The matric potential may affect seed swelling during germination and radicle or hypocotyl elongation, by its direct contribution to the effective soil mechanical stress. Under normal conditions the stresses induced in the soil are too small to confine seeds, to impair their swelling and germination, or to restrict seedling development. But poor germination observed in or next to compacted soils, or when seeds are entrapped in shrinking soil, can be attributed to greater mechanical constraints imposed by the soil, which are partly due to the matric potential.

Seed-Water Potential

The seed-water potential results from the osmotic water potential derived from the composition of the seed cells and the concentration of their constituents, and the turgor water potential component derived from the cell membranes and elasticity of the wall structure. In dry seeds the water potential is very low (negative in relation to free water), but as the seed imbibes water its water potential increases and becomes less negative. By changing the concentrations of their constituents and by modifying the activity and selectivity of their membranes, the seed cells can regulate their water potential relative to that of the soil in contact with the seed. These changes require energy inputs that develop from respiration and use of storage material. The lower the soil-water potential becomes, the greater effort and the faster the rate of depletion of the material stored in the seed for growth and establishment. Consequently, germination rate and final stand establishment will be impaired.

Seed and Soil Conductivities and Diffusivities to Water

Reported values of seed diffusivity to water (1.5×10^{-5} to $1.6 \times 10^{3}\,\mathrm{m^2\,day^{-1}}$) are lower than those reported for soils (4.0×10^{4} to $5 \times 10^{7}\,\mathrm{m^2\,day^{-1}}$) by several orders of magnitude. These observations suggest that under moist or wet soil conditions the rates of water uptake by seeds are not limited by soil-water movement rates toward them. Observed and calculated rates of water movement toward seeds indicate that for seeds that maintain lower water potentials than those in the soil, the soil could provide water to the seeds faster than the experimentally observed rates. These calculations strongly suggest that water uptake is controlled by the low permeability of the seed coat as well as the low seed diffusivity to water.

Seed-Coat Diffusivity to Water

In general, seed coats are nonuniform in shape and roughness, and in their ability to transmit water. Most seed coats are semi- or impermeable to water but they may have an opening (micropyle, hilum, chalaza). The few measured values of diffusivity to water of permeable, saturated seed coats (ranging from 9×10^{-2} to $3 \times 10^{2}\,\mathrm{m^2\,day^{-1}}$) were found to be equal to or lower than those for the wet seed bulk (2.4×10^{2} to $1.6 \times 10^{3}\,\mathrm{m^2\,day^{-1}}$). These values of seed-coat

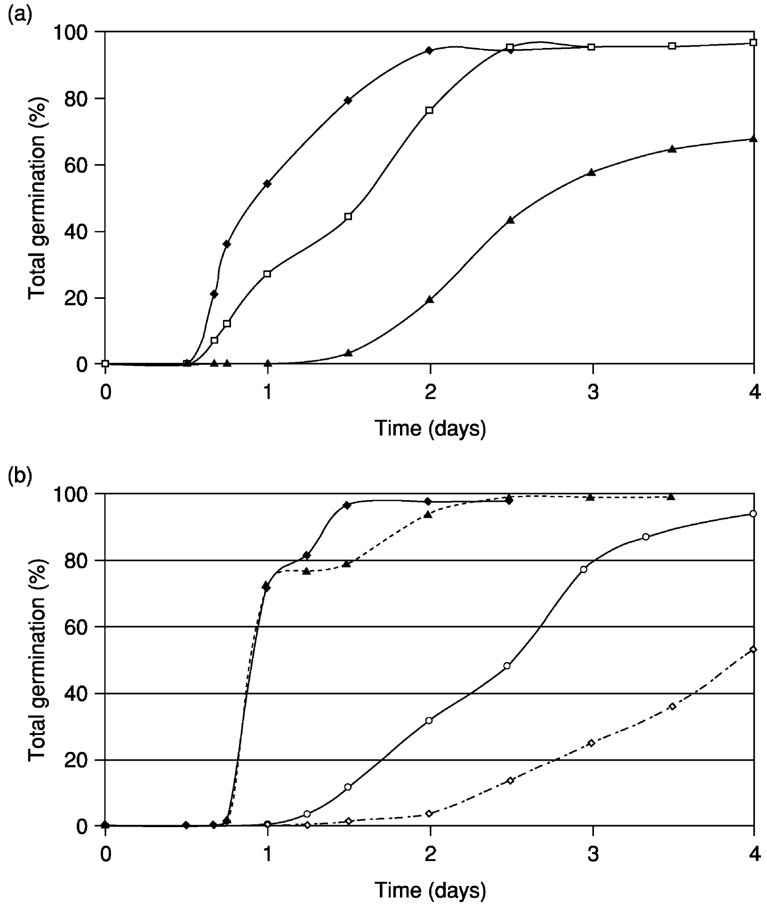


Figure 2 (a) Total germination of chickpea seeds as a function of time and matric potential: rhomboids, 100 J kg^{-1}; squares 500 J kg^{-1}; triangles 1000 J kg^{-1}. Seeds imbedded in aggregates of 0.25–0.50 mm diameter. (b) Total germination of chickpea seeds as a function of time and osmotic soil-water potential: full rhomboids 100 J kg^{-1}; triangles 330 J kg^{-1}; circles 1000 J kg^{-1}; and open rhomboids 2000 J kg^{-1}. Seeds imbedded in osmotic solutions.

diffusivity to water may affect water uptake rates and if there is also decreasing soil–seed interface contact the uptake rates could be even more strongly affected.

Seed-Coat and Seed–Soil Interface Geometrical Configuration

The geometrical configuration of the seed–soil interface zone depends on: the seed-coat surface composition and roughness; the ratio of seed size to sizes of the soil structural units in contact with it; and their spatial arrangement in the contact zone with the seed. The smaller the soil aggregates are, relative to the seed, the greater are the number of contact points and the total contact area. The contact area increases further when those contact points are wetted. Seeds with mucilaginous cover (e.g., like those of some species of the flax family) may achieve complete contact with the soil when they swell while being wetted. Hairy seeds may fail to form any direct seed–soil contact until their hairs either collapse or decay (due to the void found between the seed coat and the soil water by the mat of hairs).

Impedance to Water Flow Across the Seed–Soil Interface

The hydraulic properties of the seed–soil contact zone with respect to water vary during imbibition, because the water content at and next to the contact zone diminishes with water uptake. Since the seed–soil contact zone configuration is difficult to define and its water retention and conduction properties cannot be determined, a compound parameter (impedance to water flow) is used to describe water transfer across the contact zone. The seed–soil contact impedance to flow increases with decreasing water content in the seed–soil contact zone, seed–soil contact area, soil conductivity to water, and coarseness of soil texture and/or structure. The greater the impedance to water flow, the more restricted is the water flow from the soil toward and across the seed–soil contact interface and, consequently, the lower are the imbibition and water uptake

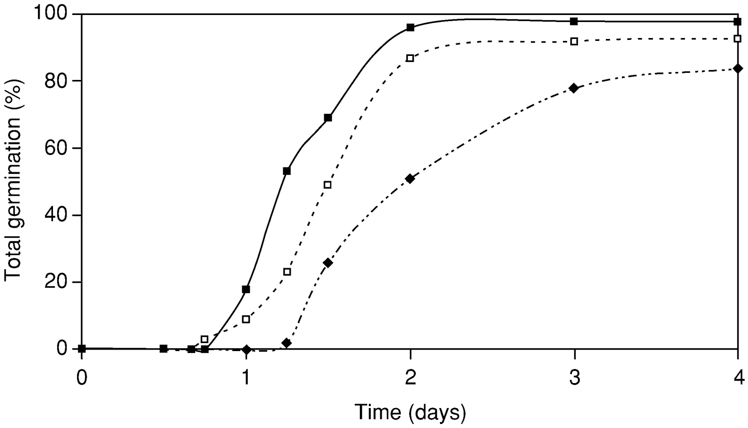


Figure 3 Total germination of chickpea seeds as a function of time and relative wetted seed surface area: filled squares 17%; open squares 6%; and rhomboids 3%. Matric potential range -10 to $-33\,\mathrm{J\,kg^{-1}}$.

rates (Figure 3). Most seeds swell during imbibition, which results in concurrent improvement of the seed–soil contact and increase in the soil mechanical stress on the seed. These changes may affect imbibition and final germination, especially in dense and compacted soils. In saline soils, seed–soil contact impedance effects may be partly obscured by the increase in the soil water osmotic potential caused by accumulation of salts excluded from entering through active membranes of the seed-surface cells during imbibition.

Seedling Emergence and Establishment

The germinated seed extends its radicle or seminal rootlets deeper into the soil to adsorb the water needed for the extension of the rootlets downward, and of the hypocotyl toward the soil surface. Water in ample quantities is needed by the rootlet and hypocotyl, to maintain the pressure required for their elongation within the soil and to overcome the weight of the overlying soil and the tensile strength of the soil surface layer, in order to emerge through it. The emerging seedling starts to transpire and to photosynthesize, and under a favorable soil-water regime, its radicle or roots elongate and start to explore the deeper soil layers in order to draw enough water to sustain the transpirational demands. It is obvious how difficult seed germination, emergence, and establishment become under arid conditions, under which the soil surface is wetted infrequently, the evaporative demands are high, and often the soils tend to slake and form a surface crust. Under such circumstances the germinating seeds and seedlings must cope with scarce and rapidly diminishing supplies of available soil water. Furthermore, soil-water depletion is too rapid either for the seedlings to extend their radicle or rootlets down into deeper, wetter soil layers where water is still available or for the seedling to break through the rapidly hardening soil crust.

Future Research

Timely, fast, and uniform seed germination, emergence, and seedling establishment are as crucial for successful crops as they are for survival of wild plants. Each seed reacts individually to the varying conditions prevailing in its microenvironment. The various aspects of seed germination and the interactions between seeds and the various environmental factors affecting their germination have been discussed above. The information presented, based on many studies carried out to resolve the complex system of seed germination and seedling establishment, is still incomplete and does not furnish a full understanding of the processes and interactions between germinating seeds and their microenvironment.

List of Technical Nomenclature

Critical hydration level ($\mathrm{cm^3\,cm^{-3}}$)	The minimal seed volumetric water content below which the seed will not germinate
Critical water potential (kPa)	A given soil-water potential value below which the seed will not germinate
Effective soil mechanical stress (kPa)	The stress transmitted through the soil by intergranular and water menisci in the soil voids
Envelope soil-water pressure (kPa)	The magnitude of change in the total soil-water potential at a given position caused by the mechanical stress imposed by the soil layers above it
Growth	The last phase of seed germination, during which the radicles and rootlets are protruding and elongating into the soil

Imbibition	A first phase of seed germination, during which the seed absorbs water and its tissues are rehydrated
Impedance to flow, soil–seed interface impedance to water flow (day^{-1})	A condition which hinders the movement of water through the seed-soil contact zone under the influence of water potential gradient
Matric potential (kPa)	The difference between the water potential of source pool of soil solution at a given elevation and air pressure and a pool identical to the source but at the elevation under consideration (above water table)
Mechanical impedance	Soil strength conditions which hinder or inhibit the devleopment and growth of plant organs in the soil
Osmotic soil-water potential (kPa)	The magnitude of change in soil-water potential due to differences in the chemical composition of the soil solution related to free pure water at the same elevation and air pressure
Seed diffusivity to water ($\text{cm}^2\,\text{day}^{-1}$)	The flux of water per unit gradient of volumetric water content in the seed
Seed germination	A general term describing the sequence of processes involved in starting metabolic activities and leading to the initiation of growth in the quiescent embryo in the seed
Seedling emergence	A term stating that a growing seedling has broken through the soil surface
Seedling establishment	A term describing the stable development and growth of a seedling that has recently emerged
Seed's volumetric water content ($\text{cm}^3\,\text{cm}^{-3}$)	The volume of water per unit bulk volume of the seed
Soil conductivity to water ($\text{cm}\,\text{day}^{-1}$)	The flux of water per unit gradient of matric potential
Soil diffusivity to water ($\text{cm}^2\,\text{day}^{-1}$)	The flux of water per unit of gradient of volumetric water content in the soil
Soil strength	A transient, localized soil property which is a combined measure of a soil to resist deformation under external or internal stress imposed on it
Soil structure	The spatial arrangement of soil primary particles into secondary particles or units
Soil texture	The relative proportions of the various size classed separates in the soil
Soil-water potential (kPa)	The amount of work that must be done in order to transport reversibly and isothermally an infinitesimal quantity of pure water from a specified source to a specified destination
Transition	The second phase in seed germination

See also: **Crusts:** Structural; **Cultivation and Tillage**; **Nutrient Availability**; **Plant–Water Relations**; **Rhizosphere**; **Root Architecture and Growth**; **Structure**

Further Reading

Arndt W (1965) The nature of the mechanical impedance to seedlings by soil surface seals. *Australian Journal of Soil Research* 3: 45–54.

Benesch-Arnold RL and Sanchez RA (1995) Modeling weed seed germination. In: Kigel J and Galili G (eds) *Seed Germination and Development*, pp. 545–566. New York: Marcel Dekker.

Bewley JD and Black M (1982) *Physiology and Biochemistry of Seeds*. Berlin, Germany: Springer-Verlag.

Bradford KL (1995) Water relations in seed germination. In: Kigel J and Galili J (eds) *Seed Development and Germination*, pp. 351–396. New York: Marcel Dekker.

Corbineau F and Come D (1995) Control of seed germination by dormancy and by the gaseous environment. In: Kigel J and Galili J (eds) *Seed Development and Germination*, pp. 397–424. New York: Marcel Dekker.

Hadas A (1982) Seed soil contact and germination. In: Khan AA (ed.) *The Physiology and Biochemistry of Seed Development, Dormancy and Germination*, pp. 507–527. Amsterdam, the Netherlands: Elsevier.

Hadas A (1997) Soil tilth – the desired soil structural state obtained through proper soil fragmentation and reorientation processes. *Soil Tillage Research* 43: 7–40.

Hadas A and Stibbe E (1977) Soil crusting and emergence of wheat seedlings. *Agronomy Journal* 69: 547–550.

Koller D and Hadas A (1982) Water relations in the germination of seeds. In: Lange OL, Nobel PS, Osmond CB, and Zigler H (eds) *Encyclopedia of Plant Physiology*, vol. 12B, pp. 410–431. Berlin, Germany: Springer–Verlag.

Kutilek M and Nielsen DR (1994) *Soil Hydrology*. Cremlingen-Destedt, Germany: Catena-Verlag.

Marshal TJ, Holmes JW, and Rose CW (1996) *Soil Physics*. Cambridge, UK: Cambridge University Press.

Mayer AM and Poljakoff-Mayber A (1989) *The Germination of Seeds*, 4th edn. London, UK: Pergamon Press.

Van Duin RHA (1956) *On the Influence of Heat, Diffusion of Air and Infiltration of Water in the Soil*, Versl. Landb. K. Onderz. 62, pp. 1–82. Pudoc: Wageningen, the Netherlands.

Global Warming *See* **Carbon Emissions and Sequestration**; **Climate Change Impacts**; **Greenhouse Gas Emissions**

GRASSLAND SOILS

J A Mason, University of Wisconsin–Madison, Madison, WI, USA
C W Zanner, University of Nebraska–Lincoln, Lincoln, NE, USA

Introduction

Soil formation in grasslands is strongly influenced by the climatic conditions under which grassland vegetation predominates as well as the distinctive characteristics of grassland ecosystems. In most grasslands, frequent soil-moisture deficits limit the rate of mineral weathering and often lead to secondary carbonate mineral accumulation in lower soil horizons. In grassland ecosystems, both the relative abundance of belowground biomass and active bioturbation lead to thick, dark, organic matter-rich A horizons. In well-drained grassland soils, organic matter content increases with increasing effective moisture and decreasing mean annual soil temperature; depth to secondary carbonates decreases with decreasing effective moisture. Topographically controlled local variation in soil hydrology can produce similar gradients within the length of a single hillslope.

Global Distribution of Grasslands and Grassland Soils

Grassland ecosystems presently (*c.* 2003) occupy between 14% and 26% of the earth's land surface. These estimates do not include large areas of former grasslands recently brought into cultivation, where the present soils formed as part of natural grassland ecosystems. The range in estimated extent in part reflects varying classification of natural plant communities that include both grasses and scattered shrubs and trees. Soil development beneath all of these communities may be strongly influenced by grasses.

From a global perspective, the distribution of grassland is largely a function of climate, particularly climatic factors that influence soil moisture availability. Effective moisture (annual precipitation minus annual potential evapotranspiration) is a useful index of those factors, although it neglects important seasonal variations in moisture supply. At the broad, continental scale, grasslands occupy regions of intermediate effective moisture, between the humid regions that support forest and arid regions dominated by desert shrub communities. The soil-water balance in grasslands varies widely, but, in the most extensive grassland regions such as the steppes of Russia and Central Asia or the North American Great Plains, potential evapotranspiration frequently exceeds precipitation both on an annual basis and during the growing season. Large interannual variability in precipitation is common and droughts can cause drastic reduction in plant productivity as well as changes in species composition. Geologic records also indicate longer-term shifts between wetter and drier climates during the Quaternary (the last 1 700 000 years); thus, most grassland soils have formed under a range of climatic conditions and plant communities.

Wind-blown sand and loess are probably the most common geologic materials underlying grasslands, mainly as a consequence of the climatic setting. Eolian sand and dust were generated locally within grasslands during past periods of dry climate, or are derived from neighboring arid regions. In either case, the wind-blown sediment can subsequently be stabilized by grassland vegetation and preserved over long time periods. Grassland ecosystems also thrive on a wide range of other geologic materials, from weathered bedrock to glacial till and alluvium.

At the local scale, factors other than regional climate become important in determining boundaries between grassland and other vegetation. Fire, ignited by lightning or by humans, is believed to have maintained grassland or savanna in regions where the climate would otherwise have allowed forest development. Recurrent fires generally favor grasses with ground-level or underground growth points over woody plants and shrubs that have most of their biomass aboveground. For example, in the 'prairie peninsula' of the midwestern USA, grasslands formerly extended far eastward into zones receiving 800–1000 mm of annual precipitation, under which forest might be expected. In this region, the natural forest–grassland boundary was often associated with obstacles to fire such as river valleys or lakes. This effect of fire has resulted in a complex pattern of natural vegetation, with isolated patches of forest surrounded by grassland or vice versa (Figure 1).

Certain soil orders are traditionally thought of as archetypal grassland soils, including Mollisols in the US Department of Agriculture's (USDA) *Soil Taxonomy*, and Chernozems and Kastanozems in the system adopted by the Food and Agriculture Organization of the UN (FAO). It is important to recognize that the global extent of these soil orders is much smaller than the extent of all soils formed under grassland or savanna ecosystems. Classification in the Mollisol order requires a mollic epipedon, one or more dark-colored, near-surface horizons that reach a minimum total thickness (generally 25 cm, though thinner mollic epipedons are allowed in some cases). Base saturation (percentage of the cation exchange capacity that is occupied by Ca^{2+}, Mg^{2+}, K^{+}, and Na^{+}) also must exceed 50% in a mollic epipedon. Subtropical or tropical grassland soils often do not qualify as Mollisols because of low organic matter content, resulting in light-colored upper horizons; some of these low-latitude soils may also be too acidic to qualify. Even at higher latitudes, large portions of most grassland regions are occupied by weakly developed soils that do not meet mollic epipedon thickness requirements (FAO: Regosols and Lithosols; USDA: Entisols and Inceptisols), particularly on eolian sand deposits and on steep slopes in a variety of parent materials. Nonetheless, Mollisols or 'Mollisol-like' soils provide the best examples of pedogenic processes specific to grasslands. Mollisols (and roughly equivalent soil orders in other systems) are found almost entirely within grasslands, or grasslands recently converted to agricultural use, whereas other soil orders that occur in grasslands can also form in a range of other ecosystems.

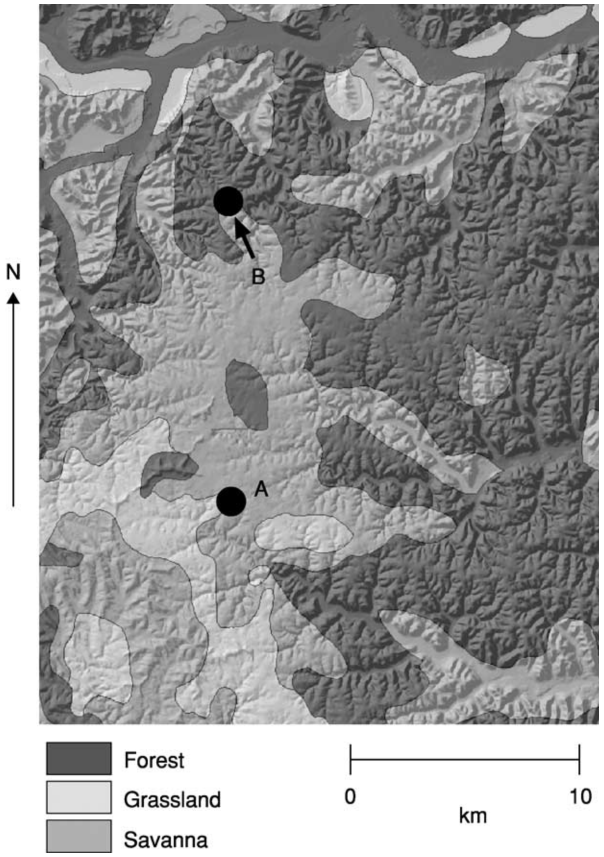


Figure 1 Complex pattern of grassland, savanna, and forest vegetation in a small area of southeastern Minnesota, USA. Vegetation types before conversion to agriculture in the 1850s to 1870s are superimposed on a shaded relief image of topography. Forest is largely confined to dissected areas where deep narrow valleys blocked rapid spread of fire. Grassland occurs on broad, gently sloping ridgetops and in some wide river valleys, where fires spread quickly. Savanna occurred in intermediate landscapes. Locations of grassland (A) and forest (B) soil profiles compared in Figure 5 are indicated. Vegetation cover is based on public land survey records, interpreted by Marschner FJ (1930) *The Original Vegetation of Minnesota* (1:500 000 scale map), US Department of Agriculture, and digitized by the Minnesota Department of Natural Resources. Topographic data are derived from US Geological Survey topographic maps.

Environment of Pedogenesis in Grasslands

The distinctive characteristics of grassland soils in part reflect the influence of the climatic conditions under which grassland vegetation predominates. The frequent soil-moisture deficits in many grassland regions inhibit silicate mineral weathering and organic matter oxidation, and favor precipitation of secondary carbonates within the soil profile.

Ecologic processes also influence grassland soil development. Net primary productivity (biomass production by photosynthesis minus biomass consumed through respiration by the photosynthesizers themselves) is relatively modest in most grasslands, relative to many forest ecosystems. The total standing biomass in grasslands is also small compared with mature forests. A larger proportion of the total biomass produced by photosynthesis is transferred to plant roots in grasslands than in forests, however. In many grasslands, underground plant biomass is greater than aboveground biomass, in some cases by a factor of 5 or more. As a consequence, the fraction of the total organic matter input that is added to the soil belowground is larger in grasslands than in forest, where most input occurs in the form of surface litter.

Consistent with the relative importance of belowground biomass, burrowing activity by animals appears to be greater in grasslands than in other ecosystems, although the difference is difficult to quantify. There are many observations of remarkably extensive burrowing by earthworms, insects (especially ants and cicadas), and rodents in North American and Eurasian grasslands. Termite-burrowing and mound construction is extensive in tropical and subtropical savannas and grasslands.

Geomorphic processes that interact with pedogenesis may also differ between grasslands and other ecosystems. As noted above, eolian sand and loess deposits are widespread in grassland regions, indicating the potential for past truncation or upbuilding of soil profiles by wind erosion and dust deposition. This potential is quite evident in grass-covered dune fields, where excavations commonly reveal multiple buried soils. The potential for reworking of soil material by slopewash is also higher in most grasslands than in forest.

Organic Matter Accumulation, Bioturbation, and Structure Development

One of the most distinctive aspects of soil formation under grassland in temperate climates was recognized in an early classic of pedology, V.V. Dokuchaev's *Russian Chernozem*. The term 'chernozem,' or 'black earth,' is an apt description of the thick, dark-colored zone of organic matter accumulation, or A horizon, in soils of relatively humid grasslands in Russia and the Ukraine. Moving southward from the chernozem zone in Russia, effective moisture decreases and A horizons become thinner and lighter-colored, reflecting lower organic carbon content. Similar trends in carbon content and A horizon color are observed along east-to-west transects of decreasing effective moisture in the grasslands of central North America (Figures 2–4) and are likely to be present in other regions. Local anomalies occur within these broad regional trends in areas where A horizons are being built upward by ongoing eolian or alluvial deposition, and soil carbon is thereby being sequestered in deeper horizons beyond the depth of most active decomposition (Figure 4b). A temperature effect independent of soil-water balance is also evident. In North America, soils in the cold grasslands of the northeastern Great Plains store more organic carbon than grassland soils farther south that receive similar effective moisture but formed under warmer temperatures. Organic carbon content also increases as temperature decreases and precipitation increases, from low- to high-elevation grasslands.

These trends with respect to climate can be interpreted using a simple model, in which total soil organic carbon content is in a steady state, with additions equal to losses. The rate of addition is a function of primary productivity, which increases with both increasing effective moisture and increasing temperature. For example, in the central Great Plains of North America, primary productivity increases by two- to threefold in an east-to-west traverse from 250 to 750 mm of annual precipitation. Each year's addition of organic matter is offset by loss through oxidation, largely through microbial activity, of a fraction of the total carbon pool. The fraction oxidized also increases with temperature and effective moisture, but not as fast as primary production. Thus, a much larger pool of soil carbon is required to maintain steady state in cool, humid grasslands than in areas that are warmer and/or drier. This model is simplistic, because the climate in most grasslands is too variable to allow a true steady state to be maintained, and various forms of organic matter are oxidized at rates differing by orders of magnitude. Most recent modeling efforts consider multiple pools of soil carbon, turning over at varying rates, but the results are still consistent with observed trends of grassland soil carbon content in relation to climate.

Soil texture also strongly influences organic matter content. Regression models for prediction of soil organic carbon content often include clay content or some other textural index as a key predictor. This effect is related to the formation of complexes containing both humus and mineral particles, which can

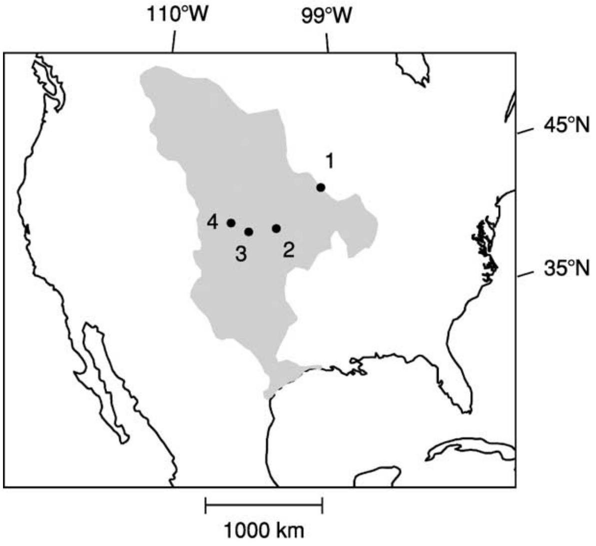


Figure 2 Locations of grassland soil profiles shown in Figure 3 and 4, within the grasslands of central North America (shaded). 1, Caledonia, Minnesota; 2, Seward, Nebraska; 3, Beaver City, Nebraska; 4, Brule, Nebraska. Boundaries of grassland region are approximate and encompass numerous outlying patches of forest and savanna, particularly near the eastern margin of grassland.

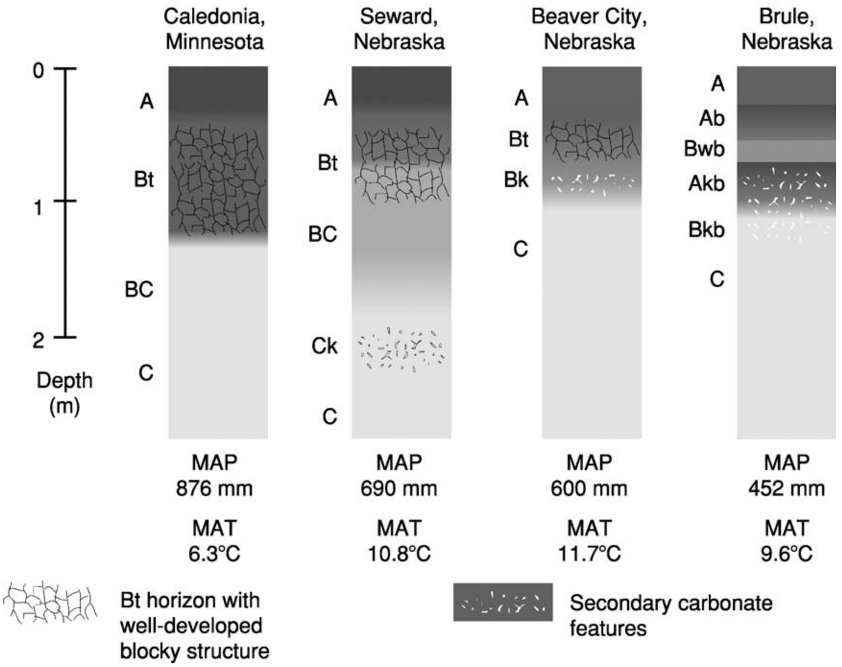


Figure 3 Schematic view of soil profiles from humid to semiarid grasslands in central North America (locations shown in Figure 2), showing typical trends associated with climatic gradients. Climatic normals indicated below each column (MAP, mean annual precipitation, 1961–90; MAT, mean annual temperature, 1961–90). The climate at each site has varied significantly over geologic time but the regional climate gradients are likely to have been maintained throughout the period of soil development. Major soil horizons are indicated. Note that ‘t’ indicates clay accumulation, ‘k’ secondary carbonates, and ‘b’ buried horizons covered by younger sediment. All profiles are formed in late Pleistocene loess deposited 25 000–12 000 years ago, except for the soil at Brule, Nebraska. The latter is formed in late Pleistocene loess at greater than 75 cm depth and in at least two increments of Holocene loess at less than 75 cm. Thickness of dark-colored upper horizons decreases westward as rainfall decreases from Caledonia to Beaver City. The Brule profile deviates from this trend because renewed loess deposition has buried the upper Ab horizon, and organic matter then accumulated in the newly deposited loess, in effect producing an overthickened A horizon. Thickness of the Bt horizon and depth to secondary carbonates also decrease with decreasing rainfall.

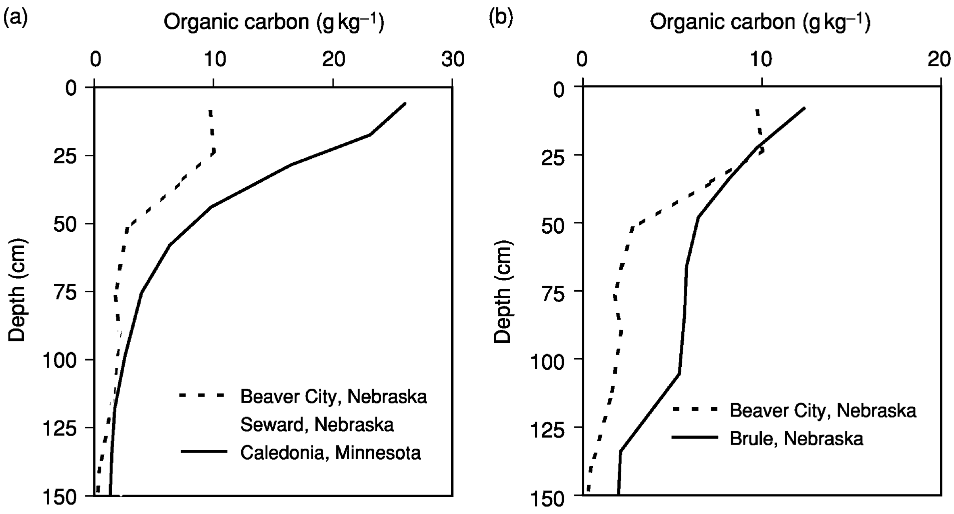


Figure 4 (a) Profiles of organic carbon content in three of the soils shown in Figure 3. Total organic carbon storage decreases with decreasing mean annual precipitation. The Caledonia profile also formed in a somewhat colder climate, which further enhances organic carbon accumulation; (b) effect of ongoing loess deposition on carbon storage. The Brule profile, where 75 cm of loess has accumulated in the past 10 000 years contains more organic carbon to a greater depth than the Beaver City profile, where mean annual precipitation is somewhat greater but loess deposition in the same time interval has been less than 40 cm.

enhance resistance to organic matter decomposition. Organic carbon associated with coarse clay and fine silt is more stable than the more rapidly turned over portion associated with fine clays. Strongly aromatic humic acids are associated with the coarser clays. Original resistance, complexation with phenolic polymers, and absorption to clay can maintain humic materials in the soil for thousands of years.

Locally, the vertical distribution of organic matter within the soil profile can change rapidly across the transition from grassland to forest. Scattered trees have limited influence on surface soil characteristics, but, as the forest becomes denser, organic matter addition to the soil is increasingly dominated by litter deposited on the soil surface rather than belowground production. Also, a portion of tree root biomass is comprised of large, suberized and lignified roots that are longer lived and less readily decomposed than fine grass roots and therefore contribute less organic matter to the soil on an annual basis. Addition primarily at the surface and limited bioturbation lead to much thinner A horizons in near-surface horizons under closed forest than under nearby grasslands. Soils with intermediate A-horizon thickness occur in the grassland–forest transition zone. Enhanced moisture in the litter layer in a shaded forest floor environment causes higher decomposition rates than in prairie and increases the production of translocatable organic acids, with the effect of reducing total carbon storage in near-surface horizons under forest (Figure 5).

Both bioturbation and formation of mineral–organic matter complexes play a role in the development of the fine, exceptionally stable structure of many A horizons in temperate grasslands. Macroscopically, soils under undisturbed natural grassland often have moderate to strong granular structure. Microscopically, these soils display a spongy fabric made up of rounded aggregates with abundant inter-aggregate pore space. Many of the most stable and well-rounded aggregates are in fact excrements of soil fauna. Others are produced through extensive burrowing activity and propagation of the fine root network of grasses. An extreme case is represented by soils that are predominantly made up of earthworm casts (Vermustolls and Vermudolls). Aggregates in undisturbed grassland soils are stabilized by mineral–organic matter complexes (particularly microbial polysaccharides interacting with clays at submicron size), fungal hyphae, and other bonding agents, and are resistant to physical disruption and slaking or dispersion in water.

Agricultural tillage breaks up soil aggregates, distributes oxygen and moisture through the plow layer, and agriculture can also add nitrogen to the system. Conversion of grasslands to agricultural use results in a reduction of the organic carbon content by 20–50% in 20–40 years. Loss of organic carbon reduces the soil's water-holding capacity and structural stability, and may reduce cation exchange capacity.

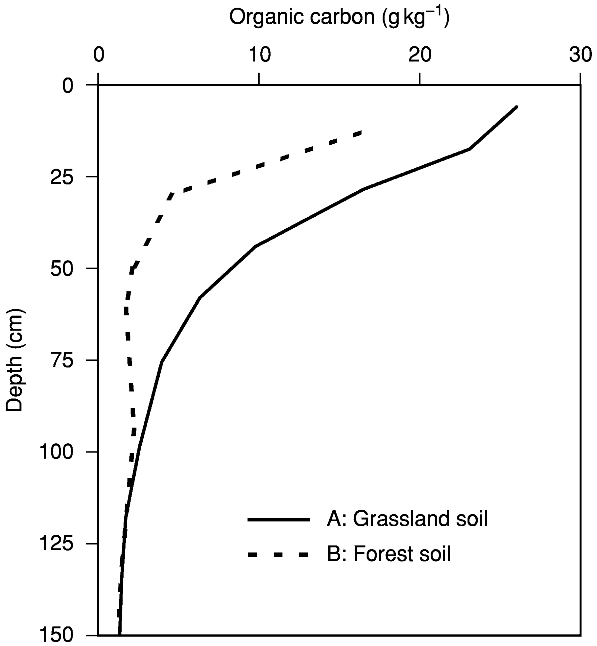


Figure 5 Contrasting organic carbon content profiles of soils in Minnesota, USA, that were under grassland or forest at the time of public land surveys in the nineteenth century. The soils formed under identical parent materials and climatic conditions and are in similar topographic settings. Locations shown in Figure 1.

Mineral Weathering, Clay Translocation, and Secondary Carbonate Mineral Formation

The typical climatic setting of grassland soils does not favor rapid silicate mineral weathering, although some studies have demonstrated that measurable chemical weathering does occur over periods of thousands to tens of thousands of years. The soil-water balance in this setting does not allow large volumes of water flux through soil; thus, the products of weathering are not effectively removed from the soil profile. Particularly in temperate grassland soils, the cation exchange complex remains dominated by Ca^{2+} and Mg^{2+}, in contrast to soils with greater throughput of water, in which these cations are depleted over time. The base saturation requirement for mollic epipedons is consistent with this observation. In the subtropical grasslands and savannas of the Sahel region in Africa, exchangeable Ca^{2+} and Mg^{2+} are also replenished by dust influx from the Sahara. A similar process may occur in other grasslands bordering desert dust sources. The predominance of Ca^{2+}, Mg^{2+}, and other base cations buffers soil pH in the circumneutral range in which silicate mineral weathering is generally slowest. Divalent cations also inhibit clay particle dispersion and translocation. In the most humid grasslands, these inhibitions on weathering and clay movement may be less effective. For example, some Mollisols in the humid Midwestern USA appear to have accumulated abundant clay in excess of the parent material clay content (Figure 6; Caledonia, Minnesota, profile).

Ecosystem characteristics as well as climate limit weathering and clay translocation in grassland soils. This is indicated by rapid changes in soil morphology across the grassland–forest transition. With increasing forest density, more prominent eluvial E horizons develop, with a more abrupt downward transition to clay-rich B horizons. Forest soils are reported to have lower pH in upper horizons (under natural vegetation), higher rates of silicate mineral weathering, and increased depletion of total phosphorus than nearby grassland soils. These contrasts can be explained by contrasting organic matter dynamics, weathering intensity, and clay mobility under the two vegetation types. Grassland organic matter decomposition tends to produce organic compounds that form strong organoclay complexes and inhibit downward translocation of base cations and clays. Under forest cover, organic acids produced as the surface residue decomposes facilitate leaching of base cations and clays. Less total organic matter is available to produce resistant organoclay complexes. These effects combine to facilitate downward translocation of clay under forest. Soils near the transition between present-day forest and prairie often show evidence of both ecosystems. This is because this boundary is not static and will shift in reaction to changes in climate.

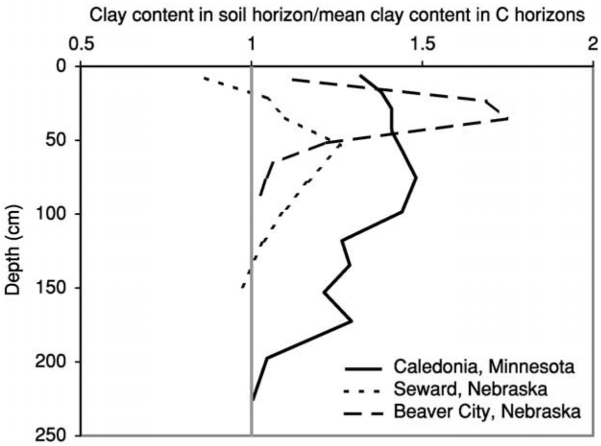


Figure 6 Clay accumulation in three of the soil profiles shown in Figure 3. Clay contents in A, B, and BC horizons are normalized by taking the ratio of actual clay content to average C-horizon clay content. Vertical bar represents a ratio of 1; values greater than 1 represent net gain of clay accumulation relative to parent material. Note prominent but shallow clay accumulation peak at the westernmost and driest site (Beaver City), and thick zone of clay accumulation at the most humid site (Caledonia).

Clay translocation in grassland soils is a controversial subject. Argillic horizons (zones of clay enrichment through translocation) are uncommon in grassland soils of Eurasia, except near the grassland–forest boundary, consistent with limited clay translocation under conditions of circumneutral pH and abundant exchangeable Ca^{2+} and Mg^{2+}. Some supposed argillic horizons in the North American Great Plains are quite prominent (Figure 6), although microscopic studies of soil thin sections provide little or no evidence of translocated clay. The German pedologist Arnt Bronger has proposed that these clay-rich horizons represent paleosols buried by loess with much lower clay content. This explanation has been received with skepticism by some North American researchers, but clear evidence for ongoing Holocene loess deposition at some sites suggests that Bronger's hypothesis is plausible (Figure 3; Brule, Nebraska, profile).

Precipitation of secondary calcite ($CaCO_3$) is nearly ubiquitous in fine-textured grassland soils of temperate regions, except in the most humid areas near the grassland–forest boundary. Secondary carbonates are less common in soils formed in sandy parent materials. In well-drained soils, the processes of secondary carbonate formation are straightforward and are clearly controlled by climate. Soil water contains dissolved Ca^{2+} derived from rainwater, dry dust deposition, or weathering of carbonate or silicate minerals. During major infiltration events, the Ca^{2+} is translocated to deeper soil horizons, but the amount of infiltrating water is insufficient to completely remove Ca^{2+} from the profile. As the soil subsequently dries, the soil water becomes saturated with respect to calcite, which precipitates (the required anion, HCO_3^-, is ubiquitous in soil water). As effective moisture decreases across a grassland region, the depth of frequent wetting decreases and secondary carbonates are concentrated at a shallower depth.

When many soils distributed across a wide range of climatic conditions are considered, a systematic decrease in depth to the upper limit of secondary carbonates with decreasing effective moisture is evident (Figure 3). At more local scales, however, depth to secondary carbonates is surprisingly variable. This variation may reflect differing soil hydraulic properties, or local topographically controlled differences in the partition of precipitation between runoff and infiltration.

Grassland Soil Hydrology and Its Influence on Pedogenesis

Closed depressions are common in many midlatitude grassland regions. These include regions that were glaciated during the most recent glacial period, where there has not been enough time for a connected drainage system to form. Closed depressions of eolian origin are common in dune fields and loess-mantled landscapes. Depressions can serve either as points of focused groundwater recharge or as discharge areas for groundwater flow systems. Some depressions alternate between these two roles on a seasonal basis.

Precipitation and snowmelt in recharge depressions is removed by both percolation and evapotranspiration (unless artificial drainage has been installed). Grassland landscapes containing closed depressions frequently illustrate the concept of a soil catena. A catena is a linked group of soils that have different characteristics because of variations in hillslope position and location within local surface and subsurface hydrologic systems. Well-drained soils on broad summits are drier and have moderately deep, dark brown surface colors reflecting more oxidizing conditions. Carbonates are leached as water infiltrates into the soil more than it runs off. Shoulder positions experience erosion and runoff, so soils are thin, and carbonates are close to the surface. Carbonate leaching depth and solum thickness increase downhill on side slopes as infiltration increases. Redoximorphic features may become more strongly expressed and closer to the surface with distance downslope, reflecting more frequent near-saturated conditions in which reducing conditions can develop. Areas low in the landscape produce more biomass and thus have greater organic inputs to the soil than upland positions. Decomposition rates are also slow in the relatively cool, wet soils in depressions or on lower slopes. The result is higher organic matter content and surface horizon colors that are blacker than those observed in the upland positions. In these wet soils, redoximorphic features can be masked by the darkness of the surface horizons, but gleyed colors (gray colors resulting from dissolution of iron oxides and hydroxides under reducing conditions) may be present at greater depth. Soils in some closed depressions display a distinct, light-colored E horizon, which is generally lacking in other grassland soils.

Grassland soils in areas of groundwater discharge often accumulate soluble salts, gypsum, and/or secondary carbonate minerals. These minerals form from ions transported to the site by groundwater flow. As water is extracted from the soil by root uptake and surface evaporation, the concentration of ions exceeds the solubility of the minerals and they precipitate, often in order of increasing solubility along the upward groundwater flow path. That is, carbonates precipitate first, followed by gypsum and soluble salts. Both the vertical and horizontal distribution of secondary minerals are highly sensitive to the local topographic and hydrologic setting. In more humid grasslands, there are extensive soils in depressions and other areas with shallow water tables that contain secondary calcite in the A horizon (Calciaquolls in USDA *Soil Taxonomy*). In other settings, there is a vertical sequence of secondary mineral formation, from calcite at greater depths through a gypsum-enriched zone to soluble salts in shallow horizons. This sequence can lead to high levels of sodium in near-surface soil horizons, either in the form of sodium salts or as Na^+ ions held on cation exchange sites on clay and organic matter. As long as the soluble salt content is high, soil structure is stable. If the upper part of the soil profile periodically experiences downward flux of infiltrated water (e.g., in a depression that is a site of either discharge or recharge depending on the season), then salts will be removed to greater depth. The Na^+-saturated clay and organic matter particles are then easily dispersed and translocated to form a dense, nearly impermeable B horizon. Such sodium-affected soils often occur in patches associated with subtle microrelief, such as shallow depressions a few meters to tens of meters in diameter.

Classification of Grassland Soils

A long-established scheme for classifying temperate grassland soils on the basis of A horizon organic matter content and color was originally developed in Russia, but was applied in the past or is still used in other parts of Eurasia and North America. In this scheme, dark-colored Chernozems occupy the humid end of the climatic range of grassland ecosystems, while lighter-colored Chestnut soils occur in drier grasslands and shrub steppes. The FAO system maintains this subdivision (Chestnut soils become Kastanozems), though with more rigorous definitions. The USDA *Soil Taxonomy* subdivides most well-drained Mollisols on the basis of climatic setting rather than intrinsic soil properties. Many Mollisols of cold climates, such as those formed under high-altitude grasslands, are included in the Cryoll suborder. In regions with Mediterranean climates (winter wet, summer dry), such as the annual grasslands of California, USA, well-drained Mollisols are classified as Xerolls. Most well-drained Mollisols in other climates are placed in the Udoll suborder in areas of greater effective moisture, and the Ustoll suborder in drier regions. The Udoll–Ustoll boundary generally does not coincide with the Chernozem–Kastanozem boundary. Many common combinations of geologic parent material and pedogenic processes occur in more than one of these climatically defined suborders. For example, grassland soils formed in loess and affected by ongoing dust deposition include both Ustolls in the Great Plains (**Figure 3**) and Xerolls of the Palouse Prairie in eastern Washington, USA.

Both the FAO and USDA systems recognize intergrade soils formed in the grassland–forest transition (e.g., Alfic Hapludolls, USDA system) and grassland soils with evidence of clay illuviation (Luvic Chernozems, FAO system). The USDA *Soil Taxonomy* includes many wet and sodium-affected grassland soils in the Mollisol order, using suborders or great groups to distinguish them, while these soils are placed in separate orders in other systems.

The emphasis on strictly climatic criteria in the USDA *Soil Taxonomy* may be ill-advised if projections of rapid climatic change in the coming century are valid. The classification of many Mollisols could change much more rapidly than any visible or measurable properties of the soils themselves.

See also: **Range Management**

Further Reading

Aandahl AR (1972) *Soils of the Great Plains. Land Use, Crops, and Grasses*. Lincoln, NE: University of Nebraska Press.

Anderson DW (1987) Pedogenesis in the grassland and adjacent forests of the Great Plains. *Advances in Agronomy* 7: 53–93.

Baxter FP and Hole FD (1967) Ant (*Formica cinerea*) pedoturbation in a prairie soil. *Soil Science Society of America Proceedings* 31: 425–428.

Bronger A (1991) Argillic horizons in modern loess soils in an Ustic soil moisture regime: comparative studies in forest-steppe and steppe areas from eastern Europe and the United States. *Advances in Soil Science* 15: 41–90.

Coupland RT (ed.) (1993) *Ecosystems of the World 8: Natural Grasslands*. Amsterdam, the Netherlands: Elsevier.

Dokuchaev VV (1883) *Russian Chernozem*. Jerusalem, Israel: Israel Program for Scientific Translation.

Fenton TE (1983) Mollisols. In: Wilding LP, Smeck NE, and Hall GF (eds) *Pedogenesis and Soil Taxonomy. II. The Soil Order*, pp. 125–164. New York: Elsevier.

Pawluk S and Bal L (1985) Micromorphology of selected mollic epipedons. In: Douglas LA and Thompson ML (eds) *Soil Micromorphology and Soil Classification*. Soil Science Society of America Special Publication No. 15. Madison, WI: Soil Science Society of America.

Weaver JE (1954) *North American Prairie*. Lincoln, NE: University of Nebraska Press.

Weaver JE and Albertson FW (1936) Effects of the great drought on the prairies of Iowa, Nebraska, and Kansas. *Ecology* 17: 567–639.

Green Manuring *See* **Cover Crops**

GREENHOUSE GAS EMISSIONS

K A Smith, University of Edinburgh, Edinburgh, UK

Introduction

The ‘greenhouse effect’ is the enhanced warming of the Earth’s surface and lower atmosphere as a result of human activity, over and above the warming due to natural processes. Absorption of infrared radiation by natural concentrations of water vapor, carbon dioxide (CO_2), and trace gases in the atmosphere is responsible for the mean surface temperature of the Earth being approximately +15°C rather than −18°C. Warming due to additional emissions of greenhouse gases is superimposed on this very large, natural warming process.

The most important individual greenhouse gas is CO_2 (*See* **Carbon Emissions and Sequestration, Carbon Cycle in Soils:** Dynamics and Management), but substantial contributions to global warming are made by other trace gases, including methane (CH_4)

and nitrous oxide (N_2O), both of which come in part from soil emissions. One kilogram of CH_4 has a warming potential 23 times greater than 1 kg of CO_2, over a 100-year period, while the warming potential of 1 kg of N_2O is nearly 300 times greater. A further contribution to global warming comes indirectly from tropospheric ozone, which is produced by light-catalyzed reactions involving nitric oxide (NO) and organic compounds emitted by vegetation. NO is a natural product emitted from soils as well as from combustion processes. Soils also act as a minor sink for methane already in the atmosphere.

Methane Sources and Sinks

The concentration of methane in the atmosphere has more than doubled since the preindustrial era, from approximately 0.7 ppm (0.7×10^{-6} mol mol^{-1}) to greater than 1.7 ppm today. The lifetime of CH_4 in the atmosphere is quite short, *c*.10 years; the main removal pathway is through oxidation by UV-created hydroxyl radicals (OH) in the atmosphere, while 5–10% is oxidized by soil microorganisms. No single source is responsible for the atmospheric increase, but soils contribute approximately 40% of total emissions (Table 1).

Soil Sources

Methane is formed in soils by the microbial breakdown of organic compounds in strictly anaerobic conditions (*See* **Anaerobic Soils**). A very low redox potential is required for this process, and CH_4 production does not begin until reduction of molecular oxygen, nitrate, iron(III), manganese(IV) and sulfate (all of which maintain the potential at higher levels) is complete. Such low-redox conditions are predominantly found in soils where prolonged waterlogging is a normal feature, e.g., natural wetlands and flooded rice fields, and beneath the soil cover of closed landfills, as well as in the sediments at the bottoms of lakes. The original identification of methane as a main component of the gas released as bubbles from natural marshes was made by Allessandro Volta in Italy in 1776. The gas is still known by the name 'marsh gas' given to it at that time.

Table 1 Estimated soil sources of methane emissions to the atmosphere

Source	*Global annual emission ($Tg\ CH_4\ year^{-1}$)*
Natural	
Wetlands	115
Termites	20
Anthropogenic	
Rice paddies	50
Landfills	40
Total from soils/soil fauna	225
Total from all sources	500–600

The organic compounds that provide the carbon for CH_4 formation in marshes, bogs, and flooded soils are released from living roots or come from the degradation of plant remains, e.g., decaying peat material and dead plant stems and leaves, in a bogland plant community, or rice straw ploughed into the soil of a paddy field prior to the establishment of the next crop. The cellulosic material in these residues is made up of hexose units, and these break down to give CH_4 and CO_2 in equimolar quantities, according to the equation:

$$C_6H_{12}O_6 \Rightarrow 3CO_2 + 3CH_4$$

The methane formed in this way can migrate to the surface and be emitted into the atmosphere by one of three different pathways (Figure 1). Diffusion can take place in solution from the point of formation in an anaerobic layer, upward to water layers containing oxygen, where much of the CH_4 is oxidized and only a fraction outgasses to the atmosphere. When sufficient gas is produced to exceed the solubility product, bubbles form in the water layer, and force their way to the surface – the process of ebullition that prompted Volta's original investigation. The speed of this process prevents any significant oxidation. The third route is via the internal structures of vascular plants. Many species that are adapted to life in flooded environments contain continuous air spaces which connect the leaves and stems to the roots. These structures, known as aerenchyma (Figure 2), have evolved to transport oxygen needed for cell division to the root meristems, but serve equally well as channels for the transport of methane from the root environment to the atmosphere.

In natural wetlands in boreal environments, CH_4 emissions have been found to vary considerably with the type of vegetation present, but the greatest differences in flux appear to be brought about by the depth to the water table. Figure 3 shows the logarithmic increase in CH_4 emissions from a range of ecosystem types in boreal North America as the measured depth to the water table diminishes from approximately 40 cm to zero. Although many investigations have been made into emissions from temperate and boreal wetlands, the degree of uncertainty about total emissions is still large. The uncertainty is greater still over emissions from tropical wetlands, because fewer studies have been made, and because there is thought to be a much bigger contribution from seasonally

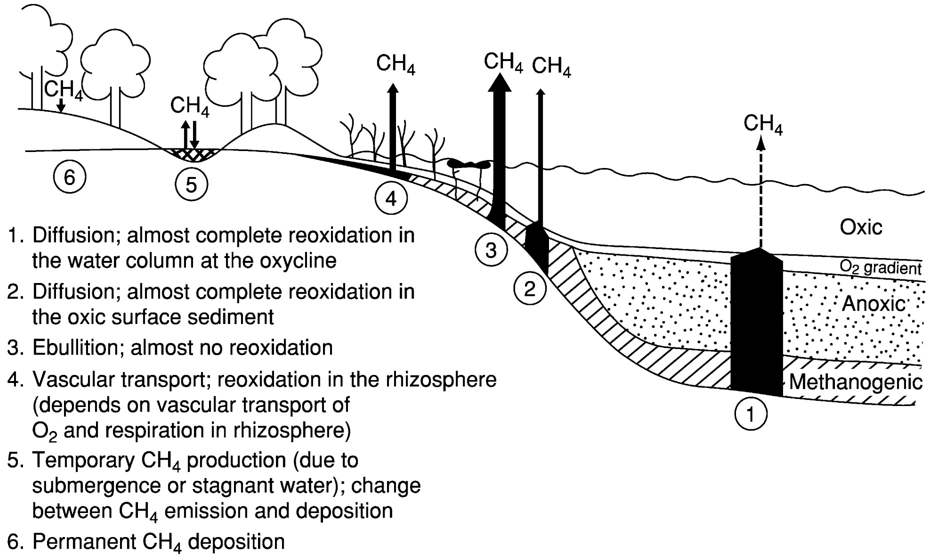


Figure 1 Transport and oxidation processes involved in methane cycling between terrestrial ecosystems and the atmosphere. (Adapted from Conrad R (1989) Control of methane production in terrestrial ecosystems. In: Andreae MO and Schimel DS (eds) *Exchange of Trace Gases between Terrestrial Ecosystems and the Atmosphere*, pp. 39–58. Chichester, UK: John Wiley & Sons, Ltd., with permission.)

flooded environments. Tropical river basins such as those of the Amazon and Parana in South America, and the Congo in Africa are prime examples of such seasonal flood plains, in which there is typically a rapid development of lush vegetation with the advent of the water, followed by decomposition under the floodwater and consequent methane production. However, the flooded areas vary considerably in extent and mean flooding depth from year to year.

In the absence of adequate experimental data, mathematical modeling of regional emissions has made an important contribution to the assessment of wetland emissions. Observed seasonal variations in the concentrations of methane in the atmosphere at different latitudes can now be simulated reasonably well by such models. Generally these models predict that more than 50% of global wetland emissions are of tropical origin.

Rice cultivation is also a major CH_4 source. The most recent estimates compiled by the Intergovernmental Panel on Climate Change have put the annual global emission from rice fields at approximately $50\,\mathrm{Tg}\,CH_4\,\mathrm{year}^{-1}$ (Table 1). The area devoted to the rice crop has expanded considerably in recent decades, and global production is predicted to rise still further in the future, to meet the needs of rising populations. This trend has undoubtedly contributed to the increase in overall CH_4 emissions observed in the last few decades. An international programme of research has been carried out to investigate the principal factors that determine the magnitude of emissions from rice. This has shown that soil type, crop fertilization, the amount of straw or organic manure incorporated into the soil before planting, and the continuity of the flooding regime, as well as climate, are all important.

In warm climates, CH_4 production also occurs in some aerobic soils, as well as in wetlands, but here the source is the bacteria that live symbiotically in the guts of some species of burrowing and mound-forming termites, and not the free-living soil microflora.

Soil Sink for Methane

The phenomenon of CH_4 oxidation by soil bacteria is general in aerated soils (Figure 1). In well-drained and therefore well-aerated soils, in the absence of termites, any CH_4 present is largely that which has diffused in from the atmosphere, where the concentration is approximately 1.7 ppm as of 2003. This methane is oxidized to CO_2 by bacteria that are adapted to living on this very low concentration of substrate. The overall global sink for atmospheric methane in soils is estimated at approximately 30 Tg of CH_4 per year; this is about 5% of total annual emissions from all sources.

Soil bulk density and water content, and their consequent effects on gas movement and penetration in the soil profile, have a major impact on the rate of oxidation of atmospheric methane in soils. Rates are highest in coarse-textured forest soils with well-developed soil structure and a permeable surface organic layer. Increasing soil acidity reduces oxidation; part of the explanation is that in acid conditions plant litter decomposes less readily and accumulates

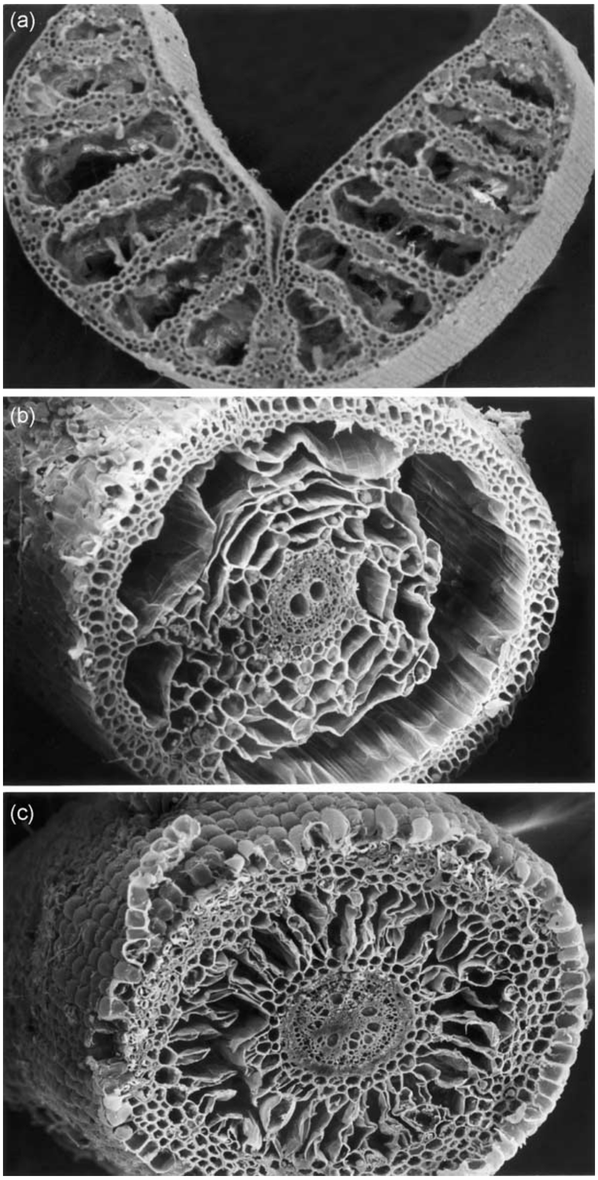


Figure 2 Scanning electron micrographs of transverse sections through the vegetative organs of some vascular wetland plants, showing the aerenchyma through which oxygen and methane can readily diffuse: (a) *Eriophorum angustifolium* leaf; (b) *Carex echinata* root (at a depth of 100–110 mm); and (c) *Molinia caerulea* root (at a depth of *c.* 150 mm). (Reproduced with permission from Thomas KL, Benstead J, Davies KL, and Lloyd D (1996) Role of wetland plants in the diurnal control of CH_4 and CO_2 fluxes in peat. *Soil Biology and Biochemistry* 28: 17–23.)

at the soil surface. This litter, particularly the leaves of broadleaved deciduous trees, acts as a diffusion barrier to gas exchange between the soil and the atmosphere. The effect of soil temperature on oxidation rate is small, and this is attributed to limitations imposed by the supply of available CH_4 substrate, due to the combined effects of diffusion resistance and low atmospheric concentration.

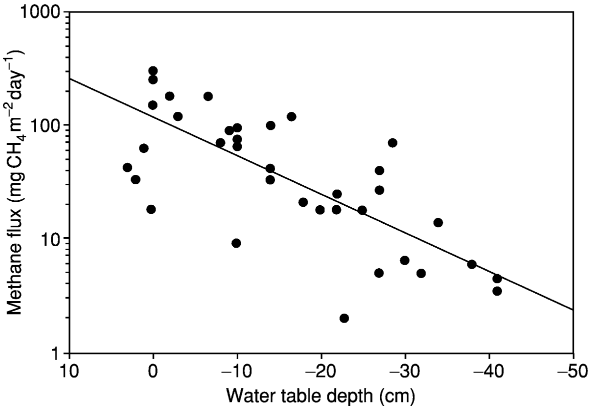


Figure 3 Relationship between methane fluxes from natural wetlands in boreal North America and water table depth (negative values on *x*-axis indicate water tables below the surface). Data compiled from six different studies. (Adapted from Liblik LK, Moore TR, Buber JL, and Robinson SD (1997) Methane emissions from wetlands in the zone of discontinuous permafrost: Fort Simpson, northwest Jemtones, Canada. *Global Biogeochemical Cycles* 11: 485–494.)

Conversion of natural soils to agricultural use reduces methane oxidation rates by approximately two-thirds. Additions of nitrogenous fertilizers also decrease oxidation rates. Recovery of oxidation rate from land-use change or fertilization is slow; it may take 100 years or more to return to predisturbance rates. The reasons for this unusually long period of adjustment are unknown.

In addition to the oxidation of atmospheric CH_4 in soils, there is also oxidation of a fraction (perhaps 20–30%) of the CH_4 produced in soils, before it escapes to the atmosphere. The methanotrophic bacteria responsible live in, for example, the thin, aerated sheath of soil that exists around the roots of vascular plants growing in flooded conditions, and in intercellular spaces within plant root systems. They also occur in aerobic hummocks on the tops of bogs, as well as in aerated cover soils above anaerobic refuse layers in landfills. In such environments, average CH_4 concentrations may be very much higher than those resulting from gas diffusion from the atmosphere, and different mixed populations of bacteria, capable of assimilating CH_4 at a faster rate than those adapted to atmospheric concentrations, become established.

Efforts to Reduce Emissions

Attempts are being made to reduce net CH_4 emissions from two major sources: sanitary landfills and flooded rice fields. Paradoxically, the creation of closed landfills for municipal wastes to prevent other problems associated with unconfined piles of refuse has brought about a major increase in the amounts of CH_4 produced. This is because the soil caps (typically 1–2 m thick) exclude atmospheric oxygen, thus allowing the landfills to become anaerobic. However, the general recognition of the undesirable contribution of these sites to global warming has led to the development of effective engineering solutions for application to large modern landfills. These range from the venting and burning on-site of the CH_4 (converting it to CO_2, which is less than one-twentieth as potent a greenhouse gas), to its collection and use in energy production, thus gaining an additional benefit through reducing the use of fossil fuel for that purpose. For many older and smaller landfills, however, there is little scope for economic installation of CH_4 collection and recovery equipment. For such environments the more effective option is the promotion of CH_4 oxidation in the cover soil. The oxidation is optimal in cover soils of intermediate textures and moderate water contents, which permit a reasonable rate of diffusion through the soil matrix and thus allow oxidation to take place, but avoid the development of cracks through which gas may be vented rapidly to the surface.

There is potential for reducing substantially the average CH_4 emissions from traditionally flooded paddy fields. This involves introducing such variations in agronomic practice as delayed flooding at the start of the season; temporary drainage of the fields in midseason; and avoiding the decomposition under flooded conditions of large amounts of organic additives such as straw and animal manure. Such changes to traditional methods can be made without detriment to crop yields and sometimes bring with them other agronomic or cost benefits – at worst they can usually be introduced without increasing the cost or difficulty of farming operations. However, persuading farmers to use such new methods takes much advisory effort, and the rate of adoption varies greatly among different societies, unless incentives are introduced.

Nitrous Oxide

The natural cycle of N_2O in the environment involves microbial and chemical production in soils and waters, and ultimate decomposition in the stratosphere. Approximately 65% of all atmospheric emissions of N_2O are from soils. Ice-core data show that the concentration of N_2O in the atmosphere in the preindustrial era was approximately 275 ppb ($275 \times 10^{-9}\,\mathrm{mol\,mol}^{-1}$), but this had increased to about 311 ppb in the early 1990s. This increase is attributed to increased anthropogenic emissions, through increased production and use of N fertilizers, tropical land conversion from forest to agriculture, increased biomass burning, etc. The extra emissions have not been accompanied by any corresponding increases in the sinks for N_2O. The dominant sink is the reaction with ozone in the stratosphere to produce NO, which was the process that first generated the environmental concern about N_2O because of its potential for increasing UV radiation at the Earth's surface. In addition, a small soil sink has been suggested, but this has not been satisfactorily quantified.

N_2O and NO are produced in soils mainly by two contrasting microbial processes, autotrophic nitrification and heterotrophic denitrification. Nitrification is the microbial oxidation of ammonium (NH_4^+) to nitrite (NO_2^-) and thence to nitrate (NO_3^-). The most studied pathway is that involving autotrophic bacteria of the genera *Nitrosomonas* and *Nitrobacter*. However, heterotrophic microorganisms, mainly fungi, may be the most important nitrifiers in some circumstances, e.g., in acid forest soils.

The nitrification process is fundamentally an aerobic one, for which the presence of molecular oxygen (O_2) is essential. However, when the supply of O_2 is limiting, the nitrifying bacteria are able to use NO_2^- as an electron acceptor and reduce it to NO and N_2O:

$$\begin{array}{c} \mathrm{NO},\ \mathrm{N_2O}\ (\text{emission}) \\ \uparrow \\ \mathrm{NH_4^+} \Rightarrow \mathrm{NO} \Rightarrow \mathrm{NO_2^-} \Rightarrow \mathrm{NO_3^-} \end{array}$$

Rates of NO production by nitrification are much greater than the corresponding rates of N_2O production. The ratio of gaseous products to nitrate produced during nitrification increases as the partial pressure of O_2 decreases, but the efficiency of NH_4^+ oxidation also decreases. This makes it difficult to estimate the effect of changes in soil aeration on the actual amount of N_2O produced. Nonetheless, in conditions in which nitrification is able to proceed, even at a reduced rate, gas diffusion rates within the soil are likely to be sufficient to permit most of the N_2O (and NO) produced to escape to the atmosphere.

The optimum temperature range for nitrification in soils is approximately 25–35°C, although nitrifiers in soils of different climatic regions may have different optima as a result of adaptation.

The other microbial process producing N_2O is denitrification, involving the reduction of NO_3^-, in the absence of oxygen. The necessary conditions for the process occur when respiratory O_2 consumption in the soil by plant roots and soil microorganisms exceeds the rate of replenishment by diffusion from the atmosphere. Anaerobic zones within the soil profile ensue, and within these zones nitrate is the chemical species that most readily acts as an electron acceptor, and so it becomes reduced by a succession of enzymes, to nitrite, NO, N_2O, and finally molecular nitrogen, N_2:

$$NO_3^- \Rightarrow NO_2^- \Rightarrow NO \Rightarrow \overset{\overset{N_2O\ (\text{emission})}{\uparrow}}{N_2O} \Rightarrow N_2$$

The fraction of the total gaseous products of denitrification which is actually emitted to the atmosphere as N_2O depends heavily on soil structure and soil wetness. On the one hand, a very small anaerobic microsite may form in an otherwise well aerated soil profile, caused by a localized region of high respiration, and any N_2O formed in the microsite will have a high probability of escaping before being reduced to N_2. On the other hand, the soil may contain large, virtually saturated (and anaerobic) clods, into which NO_3^- ions may diffuse in solution. Any N_2O produced well within the clod can only escape after diffusing in solution to the surface, and is much more likely to be reduced to N_2 before this occurs.

The denitrification pathway to N_2O has been established as having to go via NO, yet NO is virtually absent from gaseous emissions in wet soils where anaerobic conditions occur. The accepted explanation is that NO has a much greater affinity than N_2O for enzyme binding sites, and consequently, after its formation and release into the soil, NO is much more readily readsorbed by the denitrifying organisms and further reduced. Thus NO emission from denitrifying soils is usually only detected in highly artificial experimental conditions, e.g., when the NO is swept out of a soil by a stream of inert gas (N_2 or helium). Figure 4 shows the general impact of soil wetness on the mean ratio of NO to N_2O emitted from soils. In spite of the characteristic variability associated with observations of this type, there is a clear downward trend in the ratio with increasing soil water-filled pore space (WFPS), with the overall mean falling from near 10 (log (ratio) = 1) at 50% WFPS, to less than unity at greater than 70%.

The only other process of any significance producing N_2O in soils appears to be chemodenitrification. This involves chemical reactions between NO_2^- and organic compounds such as amines under acid conditions, to form N_2, NO, and N_2O, and reduction of NO_3^- and NO_2^- by ions such as Fe^{2+} in subsoils.

The soils of humid tropical forests are the largest natural terrestrial source of N_2O emissions to the atmosphere: estimated at 3 Tg of N_2O-N per year (Table 2). In these soils, nitrogen is more rapidly

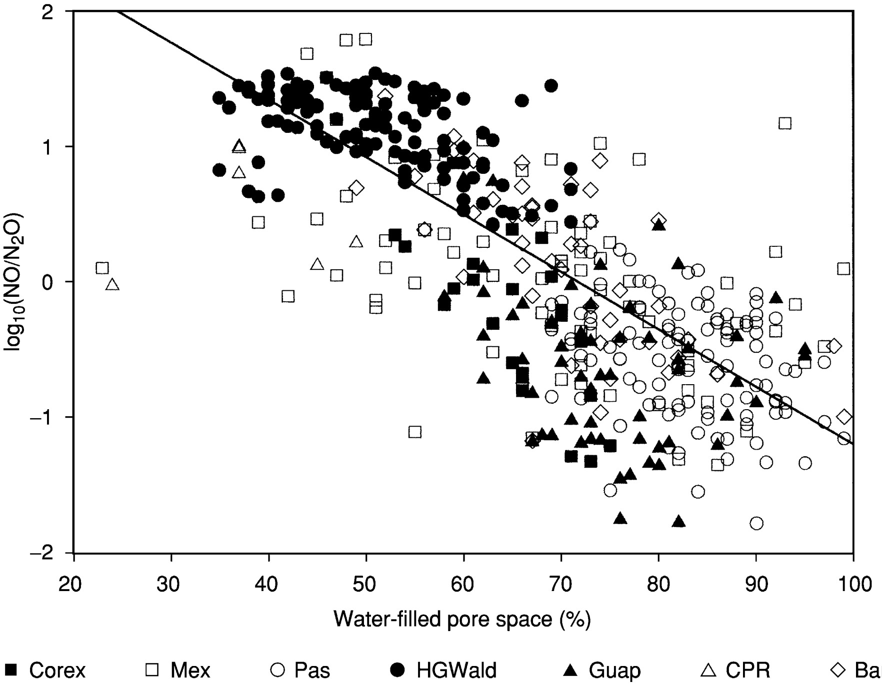


Figure 4 Relationship between soil water-filled pore space and the ratio of NO emitted from soils (g N ha^{-1} day^{-1}) to N_2O emitted (g N ha^{-1} day^{-1}), in seven studies in North and Central America and Germany. Sites: Corex, Pas, GUAP, Ban, all Costa Rica; Mex, Mexico; HGWald, Germany; CPR, Colorado, USA. Drawn from data contained in Davidson EA and Verchot LV Testing the hole-in-the-pipe model of nitric and nitrous oxide emissions from soils using the TRAGNET data base. *Global Biogeochemical Cycles* 14: 1035–1043. (Reproduced with permission from Smith KA, Ball T, Conen F *et al.* (2003) Exchange of greenhouse gases between soil and atmosphere: interactions of soil physical factors and biological processes. *European Journal of Soil Science* 54: 779–791.)

Table 2 Estimated soil sources of nitrous oxide emissions to the atmosphere

Source	*Global annual emission (Tg N_2O-N year^{-1})*
Natural	
Tropical wet forests	3.0
Tropical dry savannas	1.0
Temperate forests	1.0
Temperate grasslands	1.0
Anthropogenic	
Cultivated/fertilized soils	4.2
Total from soils	11.2
Total from all sources	17.7

cycled between organic forms and mineral (NH_4^+ and NO_3^-) forms than in temperate or boreal soils, because of the elevated temperatures and moist conditions. Thus the essential substrates for nitrification and denitrification are in more plentiful supply and the emissions are correspondingly higher. However, emissions from temperate forests, temperate grasslands and tropical dry savannas are not negligible: each is thought to contribute of the order of 1 Tg year^{-1}, giving a global total of 6 Tg N_2O-N year^{-1} from all natural soils (Table 2). The uncertainty associated with this estimate is of the order of ±50%.

Land-use change from forest to pasture or arable land has been concentrated in recent decades in the tropics. Research indicates that N_2O emissions from such newly created pastures may be several times greater than those from paired forest sites, for the first few years after deforestation, but emissions from pastures more than 20 years old are less than those from forest soil. These emissions from new agricultural land in the tropics constitute only a small fraction of the total emissions from all agricultural land across the globe. The estimated size of the latter has risen gradually, as more and more research has been carried out, and is now put at more than 4 Tg N_2O-N year^{-1} (Table 2), albeit with an uncertainty range bigger than those for the natural ecosystems.

Included within this agricultural total are emissions from land receiving mineral nitrogen fertilizers and from land receiving animal manures, either applied by farmers or directly deposited as the excreta of grazing livestock. The steady increase in consumption of nitrogenous fertilizers and in the size of livestock populations has contributed to a corresponding increase in the size of the agricultural N_2O source to its present-day level. Increases in food production are regarded as essential to meet the needs of a growing population, over the next few decades, and are likely to require increased use of N. This in turn may be expected to increase overall N_2O emissions.

Agricultural Emissions

The emissions of N_2O from agricultural soils are concentrated in those environments where there are substantial quantities of N fertilizer (and/or N in the form of organic manure) applied to the land, and where there are high animal stocking rates. Increases in soil mineral N content as a result of N application are characteristically followed by emission peaks, provided the soil is sufficiently wet (Figure 5). The largest fluxes occur when WFPS values are very high (70–90%), indicating that denitrification is the major process responsible.

On grazed land, the deposition of excreted N is spatially very variable. A small patch of soil surface may receive N as urine at a rate equivalent to some hundreds of kilograms per hectare, whereas land between patches receives none. Also, the treading of the soil surface in wet conditions by animal hooves can create localized compacted zones in which water collects and the soil becomes anaerobic. Such factors lead to the occurrence of 'hotspots' of microbial activity and N_2O emission; overall the average emission rate from grazed grassland may be two to three times that from N-fertilized grass grown as a crop to be cut for winter feed. In western Europe at least, the emissions from cut grassland are, in turn, substantially higher than emission rates from arable land cropped with wheat and barley. This is attributable to a combination of soil physical conditions and the timing of N applications. However, the emissions are very variable between sites and between successive seasons, and for any crop type several years' data are required to obtain a robust estimate of annual N_2O flux.

The response of N_2O emissions to changes in soil temperature can be complex. In conditions in which topsoils are routinely frozen for part of the winter, or are subject to successive freeze–thaw cycles, a substantial part of the total annual emission may occur within a brief period after thawing. This is true of both forest soils and arable land. The principal cause is the development of conditions that stimulate microbial activity, in particular the release of carbon from organic materials in the presence of a high soil-water content. Many research studies have reported nonlinear (exponential) increases in N_2O emission rate with temperature, with values for the Q_{10} for the process (i.e., the ratio of the rate at $T+10°$ to the rate at $T°$) ranging up to 10 or more, compared with the general range of 2–3 for most biochemical processes. An example with a Q_{10} of 5.9 is shown

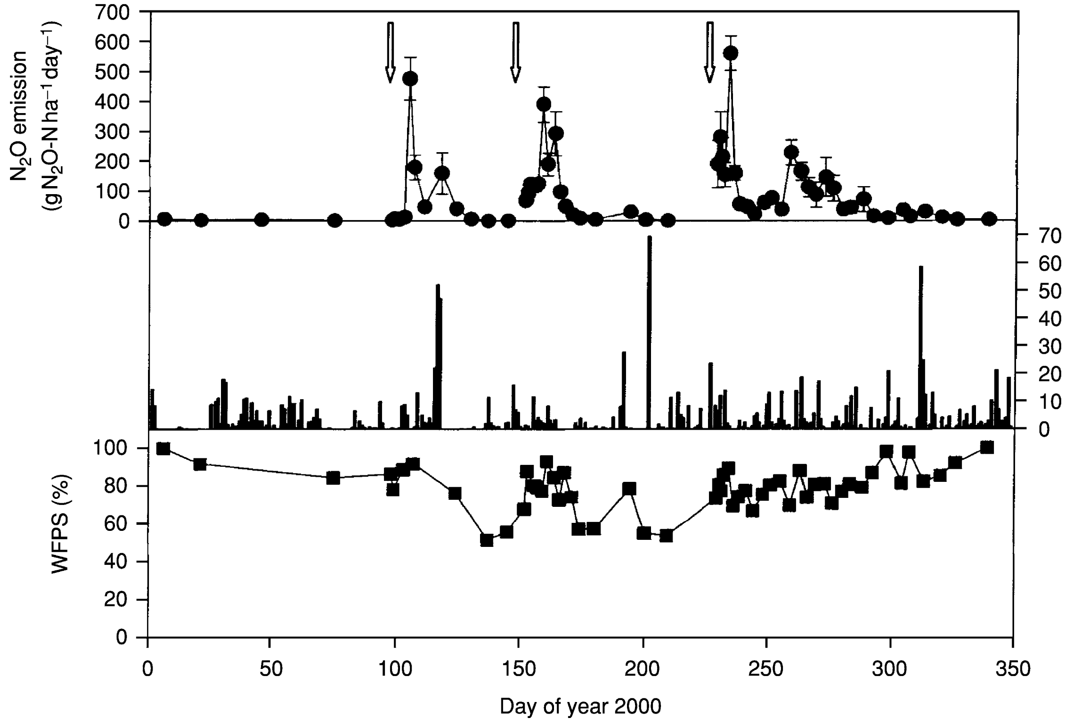


Figure 5 N_2O emissions from a glacial till soil in SE Scotland under grass, fertilized and cut for silage three times per season, together with rainfall and soil water-filled pore space data for the same site. Arrows indicate dates of fertilization (100 kg N as ammonium nitrate on each occasion). (Adapted from Dobbie KE and Smith KA (2003) Nitrous oxide emission factors for agricultural soils in Great Britain: the impact of soil water-filled pore space and other controlling variables. *Global Change Biology* 9: 204–218.)

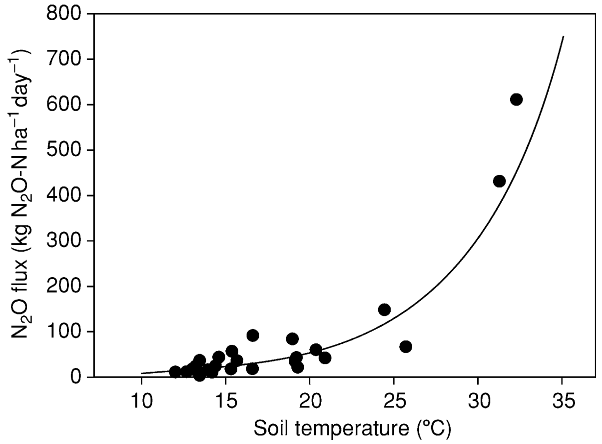


Figure 6 Exponential increase with temperature of the N_2O emission from a grassland soil fertilized with ammonium nitrate. Samples of 40 cm diameter of turf/topsoil enclosed in plastic rings and transferred from the field to the laboratory for incubation at a range of temperatures. Q_{10} for the temperature response = 5.9. (Adapted from Conen F (1994) Influence of temperature on nitrous oxide emissions from soil. MSc Dissertation. University of Edinburgh.)

in Figure 6. An explanation is provided by the theory of anaerobic zone development in soils: all N_2O production by denitrification goes on within such zones, and an increase in temperature leads both to an increase in the size of the zones (because of increased respiration depleting the oxygen supply), and in the rate of denitrification per unit anaerobic volume. Thus the overall change in the rate of N_2O production in a soil mass is the product of these two factors. An increase in soil WFPS, and the consequent decrease in O_2 diffusion rate within the soil, has the same effect, and thus N_2O production/emission increases exponentially with WFPS.

Potential for Reducing N_2O Emissions

The strong link between the amount of N used as a nutrient in agriculture and N_2O emissions means that there is some prospect of reducing emissions by improving the agronomic efficiency with which fertilizer and manure N are used. In general, the procedures recognized as helping in this direction include:

- Using soil/plant testing to determine fertilizer N needs;
- Minimizing fallow periods to limit mineral N accumulation;
- Improving the timing of and/or splitting N applications;
- Using special forms of synthetic fertilizers, such as ‘controlled release’ materials, which make the

N available for crop uptake over an extended period;

- Adding nitrification inhibitors to urea and ammonium fertilizers;
- Matching the fertilizer type to the soil water regime, e.g., by using an ammonium form in circumstances where wet soil conditions are likely to cause excessive denitrification losses from nitrate forms.

The widespread adoption of these measures could reduce total emissions by 20–25%.

See also: **Anaerobic Soils**; **Carbon Emissions and Sequestration**; **Climate Change Impacts**; **Climate Models, Role of Soil**; **Waste Disposal on Land:** Municipal

Further Reading

Andreae MO and Schimel DS (eds) (1989) *Exchange of Trace Gases Between Terrestrial Ecosystems and the Atmosphere*. Chichester, UK: John Wiley & Sons, Ltd.

Bouwman AF (ed.) (1990) *Soils and the Greenhouse Effect.* Chichester, UK: John Wiley & Sons, Ltd.

Conrad R (1996) Soil microorganisms as controllers of atmospheric trace gases (H_2, CO, CH_4, OCS, N_2O, and NO). *Microbiological Reviews* 60: 609–640.

Firestone MK (1982) Biological denitrification. In: Stevenson FJ (ed.) *Nitrogen in Agricultural Soils*, pp. 289–326. Madison, WI: American Society of Agronomy.

Granli T and Bøckman OC (1994) Nitrous oxide from agriculture. *Norwegian Journal of Agricultural Sciences* (Supplement 12) 1–128.

Houghton JT, Meiro Filho LG, Callander BA *et al.* (eds) (1996) *Climate Change 1995 – The Science of Climate Change.* Cambridge, UK: Cambridge University Press.

Houghton JT, Ding Y, Griggs DJ *et al.* (eds) (2001) *Climate Change 2001: The Scientific Basis.* Cambridge, UK: Cambridge University Press.

Prather M, Derwent R, Ehhalt D *et al.* (1995) Other trace gases and atmospheric chemistry. In: Houghton JT *et al.* (eds) *Climate Change 1994: Radiative Forcing of Climate Change and an Evaluation of the IPCC IS92 Emission Scenarios*, pp. 73–126. Cambridge, UK: Cambridge University Press.

Prinn RG (ed.) (1994) *Global Atmospheric-Biospheric Chemistry.* New York: Plenum Press.

Wassmann R, Lantin RS, and Neue H-U (eds) (2000) *Methane Emissions from Major Rice Ecosystems in Asia.* Dordrecht, the Netherlands: Kluwer.

Watson RT, Zinyowera MC, and Moss RH (eds) (1996) *Climate Change 1995. Impacts, Adaptations and Mitigation of Climate Change: Scientific-Technical Analyses.* Cambridge, UK: Cambridge University Press.

GROUNDWATER AND AQUIFERS

Y Bachmat, Hebrew University, Jerusalem, Israel

Basic Concepts

While digging a well or drilling a borehole into the ground, two main types of moisture zones might be encountered in the subsurface: An upper, unsaturated zone, in which the void space of the rock contains both liquid water and a gaseous phase (including air, water vapor, and other gases); and a lower, saturated zone in which the entire void space is filled with liquid water only. Water shows up in the well only after it has entered the saturated zone and reached a depth where the water pressure becomes equal to or higher than atmospheric. Such water is termed 'groundwater.'

Groundwater is unconfined if bounded from above by a free surface at atmospheric pressure, the 'water table.' In this case water will appear at the bottom of the well once uncovered. Groundwater is confined if bounded from above by an impervious layer at a pressure higher than atmospheric. In this case the water, once exposed, will rise to a level higher than the bottom of the well.

Although groundwater exists almost everywhere, the focus of attention is usually on groundwater found in geological formations of permeable rock or loose material, which can both contain and transmit significant amounts of water. Such formations are referred to as 'aquifers.' Particularly important among them are aquifers in which the water is renewable, in contrast to aquifers found in desert regions of the world (e.g., the Sahara Desert, the Australian Great Basin, the Sinai Peninsula), which contain fossil groundwater, namely water that has accumulated in the far past and represents a one-time, nonrenewable resource. From here on, only aquifers with renewable groundwater are considered.

Types of Aquifer

Very deep boreholes may encounter two or more layers of aquifers which are separated from each other by impervious formations or semipervious ones, called 'aquitards.' Accordingly, the saturated zone of an aquifer is said to be phreatic or unconfined wherever bounded by a water table, confined wherever bounded at top and bottom by impervious layers, leaky wherever bounded by under- or overlying aquitards, and artesian wherever groundwater is flowing out above the surface of the ground.

Groundwater Basin

Sometimes a group of several aquifers that are interconnected in the subsurface through semipermeable layers or faults may form an independent hydraulic unit. This leads to the concept of a 'groundwater basin,' which resembles a surface watershed. A groundwater basin is defined as a physiographic unit containing several connected aquifers, whose waters are flowing to a common outlet, and which is delimited in the subsurface by impermeable bondaries.

In principle, if water is added to or abstracted from an aquifer at any point, water levels and the patterns of groundwater flow in the entire basin are affected in the course of time. The same applies to the transport and concentration of a pollutant, but on a much longer time scale. The region appropriate for the control of groundwater quantity and quality should, therefore, be a groundwater basin.

A groundwater basin usually crosses the boundaries of overlying administrative areas of jurisdiction in one country and, often, also boundaries between countries. Hence, any analysis of the state of groundwater within the limits of a political or an administrative unit must take into account the effect of subsurface transport of groundwater and its constituents across the boundaries of the unit.

Prominent Features of an Aquifer

Large Areal Extent Compared with Saturated Thickness

Aquifers may stretch over hundreds and thousands of square kilometers underground, whereas the thickness of an aquifer may range from tens to hundreds of meters only. For this reason aquifers are often modeled as large but thin sheets of porous material with properties that are uniform throughout their depth and may vary only on the surface of the sheet. Such a two-dimensional representation, referred to as a hydraulic model of an aquifer, is widely used in the study of groundwater flow on a regional scale whenever water quality variations with depth are practically negligible.

Permanently yet Slowly Renewable Reservoir

Groundwater stored in an aquifer is in a state of permanent motion from areas of natural and artificial recharge to natural and artificial drainage outlets. Renewable groundwater is actually part of the hydrologic cycle, a schematic representation of which is given in Figure 1. It shows a permanent circulation of water from the oceans and land masses to the atmosphere by evaporation and transpiration and then, by condensation and precipitation, on to the surface and back to the oceans via streams or aquifers.

It is worth noting that most of the precipitation on land (in the order of 70%) is lost to the atmosphere by evapotranspiration. But most of the remainder (in the order of 25%) percolates via the unsaturated zone downward to the saturated portion of an aquifer and moves as groundwater from areas of recharge to natural discharge outlets on the surface of the ground, such as springs, marshes, streams, lakes, and oceans (see Figure 2).

It is evident that the recharge of an aquifer from precipitation and/or from other surface sources of water can only occur in areas where the unsaturated zone is permeable and the saturated portion of the aquifer is unconfined. The possibility of direct contact with surface sources of water and pollutants makes the unconfined portions of an aquifer more sensitive to changes in groundwater quality and storage than portions of the aquifer in which the groundwater is confined between impermeable layers.

The movement of groundwater within the saturated zone is usually in the order of meters per day and even per year. As a result, the residence time of a water particle in the saturated zone (i.e., the time it takes a particle to travel from its point of entry into the saturated zone to the point of natural discharge) may range from several weeks to thousands of years. The aforementioned is typical for most aquifers composed of granular material such as sand or gravel. Much faster, yet exceptional, flow velocities may prevail in aquifers of solid rock that is highly fractured or in limestone aquifers, termed 'karst,' where groundwater has dissolved the rock material to form large channels or even caverns.

Large Storage Capacity Compared with Annual Recharge/Discharge

The capacity and volume of groundwater stored in an aquifer are usually large compared with its annual recharge/discharge volume. Under such

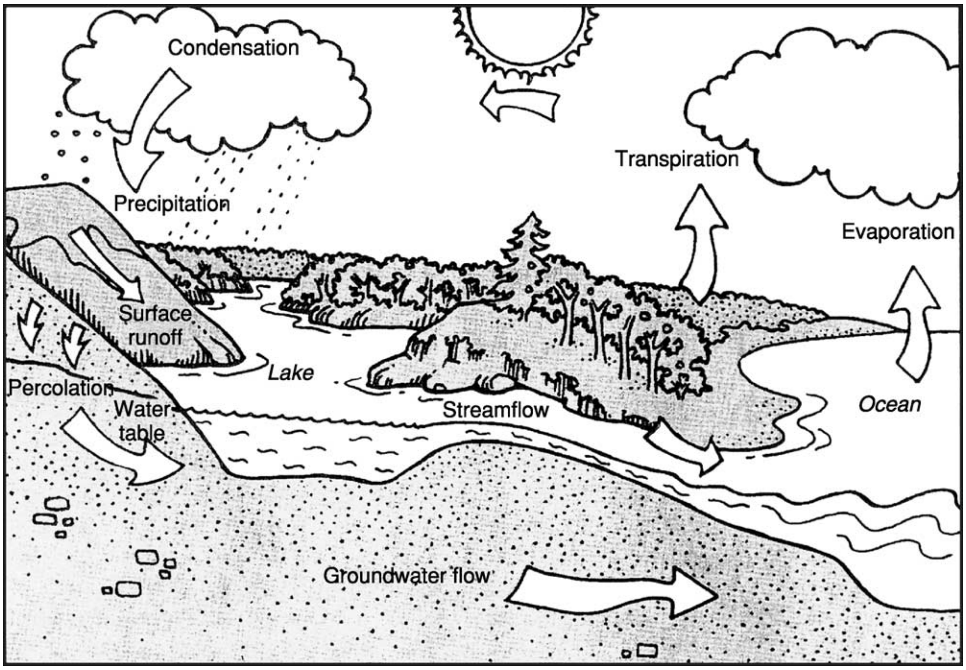


Figure 1 The hydrologic cycle. (Adapted from Bachmat Y (1994) Groundwater as part of a water system. In: Zoller U (ed.) *Groundwater Contamination and Control*. New York: Marcel Dekker, Inc. with permission.)

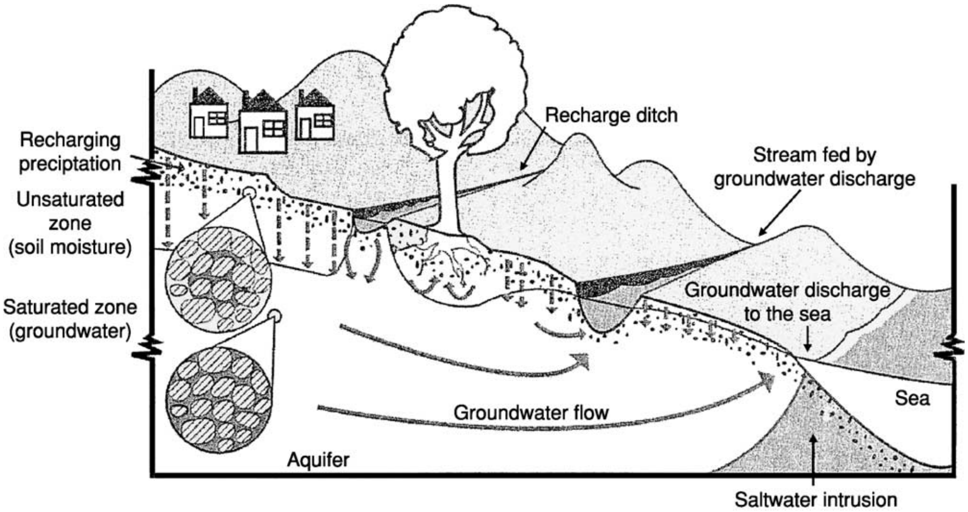


Figure 2 Groundwater flow. (Adapted from Bachmat Y (1994) Groundwater as part of a water system. In: Zoller U (ed.) *Groundwater Contamination and Control*. New York: Marcel Dekker, Inc. with permission.)

circumstances an aquifer can absorb fluctuations in annual recharge without long-term effects on the total volume or quality of the water stored in it.

Hidden Medium

The internal configuration of the void space of an aquifer is not accessible to direct observation. Moreover, samples of various types of rock material provide evidence of its complexity and diversity. Under such conditions, the spatial and temporal pattern of groundwater flow and its response to external effects can only be inferred by the conjunctive use of scientific knowledge and field records of quantities that are both relevant and observable; for example, water level observations in a piezometer.

A piezometer is a small-diameter borehole placed in the ground with an upper end open to the atmosphere and the lower end in contact with groundwater. The water level in a piezometer abve a given reference plane (e.g., the mean sea level) is the 'piezometric head.'

Time series of observations of the piezometric head at different locations and depths in the saturated zone of an aquifer are used by the theory of groundwater flow for assessing the volume and flow rates of groundwater in the aquifer in the course of time.

Functions of an Aquifer

The features of an aquifer listed above can be used to satisfy a multitude of human and environmental needs.

Easily Accessible and Durable Source of Water

The use of groundwater is facilitated by the fact that, within the domain of the areal extent of an aquifer, groundwater is found everywhere and often at small depths, ranging from tens to several hundreds of meters. Therefore, groundwater has always been an important – and in some parts of the world a major or even unique – source of fresh water supply. Groundwater is used primarily for irrigation and as a a source of drinking water in many parts of the world where surface water is nonexistent or extremely costly to develop. However, even in countries where surface water is abundant, the generally good quality of groundwater, its accessibility, and its relatively low cost are factors that stimulate an increasing level of exploitation of this resource. In the USA, for example, groundwater supplies only 25% of the fresh water used for all purposes, but approximately 50% of the population relies on groundwater as its primary source of water for drinking. In rural areas the percentage is even higher. Moreover, the use of groundwater in the USA is increasing more rapidly than the use of surface water. In Canada, 26% of the population rely on groundwater for domestic use. Two-thirds of these users live in rural areas and the rest in small municipalities, where groundwater is the primary source for their supply of water. In Israel, a semiarid country, groundwater provides three-quarters of the fresh water supply.

Subsurface Storage Reservoir

Aquifers can also be used for storing surplus amounts of water from the rainy season to the dry one, as well as from years of exessive precipitation to years of drought. Underground storage of surface water is carried out either by means of artificial recharge wells, which inject the water directly into the saturated zone of the aquifer, or by storing the water in natural depressions or artificial ponds, referred to as 'spreading grounds,' and letting the water drain through the unsaturated zone down to the groundwater. Obviously, this form of indirect recharge is practical in areas where the unsaturated zone is permeable and allows an undisturbed percolation to the underlying unconfined groundwater. In general, preference is given to artificial recharge in unconfined portions of an aquifer where groundwater has free space to accumulate on top of the water table.

An additional advantage of subsurface storage of water is the fact that it saves losses of land and losses by evaporation of water which are associated with storage of water on the surface of the ground. An example of aquifers serving as seasonal and long-term storage reservoirs of water is well illustrated by the water supply system in Israel. Israel's national water supply is based on three major elements: a lake (Lake Kinneret) that drains the water of the Jordan river basin in the northern part of the country; two major aquifers (coastal aquifer, composed of sand and sandstone, and Yarkon–Taninim aquifer, composed of karstified limestone and dolomite) in the central part of the country; and a transportation–distribution system that interconnects the above sources and links them with users. The main conduit of this system is the National Water Carrier, which transports water from the lake to the central and southern semiarid parts of the country. Both the inflow into the lake and the natural replenishment of the aquifers are highly variable, and a storage capacity of 4.5 billion cubic meters is needed for full interannual regulation of the national water supply system. Except for the lake, which has only 10% of this capacity, there are no other favorable surface storage options. Therefore, the main burden of providing the needed volume and facilities for the regulation of water supply falls on the two aquifers. Water brought by the National Water Carrier from the lake and injected into the aquifers, in the rainy winter season and in years of heavy rainfall, is repumped and transported back to the National Water Carrier to meet water needs in the summer season and in years of drought.

Facility for Improving Water Quality

The natural filtering capacity of an aquifer as a porous medium, and in particular processes of adsorption, ion exchange, dilution, and chemical and biochemical reactions that prevail in the unsaturated zone, can reduce nutrients, carbonaceous oxygen demand, certain metals, and bacteriological pathogens. The Dan Region Wastewater Reclamation Plant, located in the sand dunes of the coastal plain of Israel south of Tel Aviv, may serve as an example. This plant handles at present an annual amount of nearly 100 million cubic meters of municipal sewage collected from the greater metropolitan area of Tel Aviv. After some biological and chemical pretreatment, the partially treated effluent is brought to spreading grounds and left to percolate downwards through the unsaturated zone, which is used as a natural filter. Upon reaching the groundwater table, the water continues to move through the saturated zone, to be intercepted by a chain of pumping wells and transported to the dry south for irrigation of industrial crops.

Aquifer as a Conduit

The large spatial extent of an aquifer permits water that enters the saturated zone by natural or induced replenishment to flow long distances from the region of recharge to points of discharge or artificial withdrawal. The proximity of groundwater to individual consumers spread out over large areas saves the need for large water conveyance and distribution systems overland.

Aquifer Characteristics

Some geometric parameters related to the capacity of an aquifer to store and transmit water are:

- Porosity: volume of voids per unit elementary volume of the aquifer;
- Effective porosity: volume of interconnected voids per unit elementary volume of the aquifer. While porosity is a local measure of the total water storage capacity, effective porosity is a local measure of the aquifer's capacity to store movable water only. Hence the important role of effective porosity in the transport of groundwater and its constituents through the saturated zone of an aquifer;
- Specific storativity: volume of water added to (or released from) storage per unit volume of the aquifer per unit rise (or decline) of the piezometric head.
- Permeability (or intrinsic permeability): a characteristic of the configuration of the void space which affects its ability to transmit a fluid. The permeability is determined by three factors: (1) the effective porosity of the aquifer material; (2) the surface area of the solid matrix of the aquifer which exerts a resistance to flow. The smaller this area per unit volume of the void space, the larger the permeability of the aquifer material; (3) the tortuosity of the actual flow paths. Tortuosity is the ratio of the actual length of a flow path measured between two points to the shortest distance between the two points. The smaller the tortuosity, the larger the permeability.

Another geometric parameter of an aquifer which plays an important role in the transport of a solute is dispersivity. It represents a measure of the capacity of the aquifer to cause spreading of a solute. While permeability depends on the average value of tortuosity in the vicinity of a point in the aquifer, dispersivity depends on the variance of the tortuosity around a point.

Aquifers composed of hard rock material are rigid in the sense that they are not subject to changes of porosity due to changes in external load on the aquifer or changes of water pressure inside the saturated zone of the aquifer. The contrary is true for aquifers containing soft sediments, which are deformable, such as clay. Examples of severe environmental problems associated with the deformability of such aquifers are land subsidence due to compaction because of the drop in water pressure as a result of pumping out large amounts of water from the aquifer, or swelling because of increase in water pressure as a result of injecting large amounts of water in the confined portion of an aquifer.

Aquifers are usually heterogeneous, meaning that the properties listed above may vary from point to point within the aquifer. Aquifers may also exhibit anisotropy, meaning that the permeability at a point within the aquifer may be different in different directions. A typical example of anisotropy is permeability along a deposited layer of sediments being different from the permeability across the layer. Anisotropy is also typical for limestone aquifers where dissolution of rock material has formed large channels with preferred flow directions.

Groundwater State and Behavior

The state of groundwater in an aquifer is described by the values of a set of variables which represent physical, chemical, biological, and other properties of the water. The selection of the set of these variables, referred to as state variables, is not unique. It depends on requirements related to the intended use of the water. Common examples of state variables of groundwater are depth to groundwater, velocity, pressure, concentration of dissolved and suspended substances, and temperature.

The state of groundwater in an aquifer at a given point in time is nonuniform, namely it varies from place to place. Usually, the state of groundwater is also nonsteady, namely it varies in the course of time. The behavior of groundwater in a given period of time is the sequence of states undergone by the groundwater during that period. Knowing the behavior of groundwater in a given aquifer in the past is helpful in recognizing the factors which govern the behavior of the groundwater and in establishing quantitative cause–effect relationships between these factors and the state variables. This knowledge, in turn, can be used for predicting the behavior of the groundwater in the future. The branch of science which studies the behavior of groundwater is 'groundwater hydrology.'

Levels of Groundwater Study

The behavior of groundwater can be studied at two levels: (1) the microscopic level, which follows the movement and properties of individual water particles within the actual configuration of the void space; (2) the macroscopic level, which views the groundwater and the solid matrix as material continua occupying simultaneously the entire saturated zone. At this level of description, the value of a state variable of the groundwater at each point is an average of local values within an elementary volume around the point, referred to as representative elementary volume (REV).

The behavior of groundwater at the microscopic level is sometimes studied in research laboratories. In the field, however, observations and study of groundwater behavior are possible and of practical interest at the macroscopic level only. Therefore, and because of the ease of mathematical treatment, groundwater hydrology uses the macroscopic level of description. It is worth noting that the parameters of an aquifer presented above make sense and could be defined at the macroscopic level only.

A more simplified level of description of groundwater behavior is the hydraulic level, at which state variables and parameters are averaged over the entire depth of the saturated zone, thereby reducing the description of their variation to two spatial coordinates only. As stated earlier, the hydraulic level of description is applied wherever the variation of the values of parameters and variables of interest along the vertical are negligible in comparison with their averages.

Tracing the Movement of Groundwater

The movement of groundwater in an aquifer is described by a variable termed 'specific discharge.' It is defined at any point as the volume of water which passes per unit time through a unit area perpendicular to the direction of flow at that point. The specific discharge is also termed 'apparent velocity of the groundwater,' since it assumes that the flow takes place through the entire unit area, while the actual flow is through the area of the voids only. The specific discharge is therefore smaller than the actual average velocity by a factor equal to the effective porosity defined earlier.

Since specific discharge cannot be measured directly, attempts have been made to estimate it by indirect observations. In 1856, a French engineer, Henry Darcy, investigated the flow of water in vertical, homogeneous sand columns. He concluded that the discharge per unit cross-sectional area of a column is directly proportional to the difference in water level elevations in the inflow and exit reservoirs of the column, and inversely proportional to the length of the column. The findings of Darcy have been adapted to flow in a porous medium and given the name 'Darcy's law.'

Definitions pertinent to Darcy's law include:

- Hydraulic slope: the drop of the piezometric head per unit distance in a given direction;
- Hydraulic gradient: the vector whose direction and magnitude are those of the largest hydraulic slope. The hydraulic gradient at any point is perpendicular to the surface of constant piezometric head at that point.

Darcy's Law

The specific discharge of a homogeneous fluid (e.g., groundwater of uniform density) at any point of a homogeneous and isotropic porous medium is directly proportional to the hydraulic gradient at that point, states Darcy's law.

The coefficient of proportionality, 'hydraulic conductivity,' is directly proportional to the permeability of the medium, defined earlier. It also depends on the properties of the fluid. The heavier the fluid, the higher its conductivity; the higher the viscosity of the fluid, the lower its conductivity.

Darcy's law, wherever applicable, enables determination of the specific discharge at any point in an aquifer at any time by means of the piezometric head, which is observable. Hence the importance of monitoring and mapping the piezometric head for the assessment of groundwater flow directions and volumes.

An example is the determination of streamlines and discharges in a hydraulic model of an aquifer at any given time. According to Darcy's law, at any point in the groundwater flow domain, a line perpendicular to a contour line (i.e., an isoline) of the piezometric head is a streamline. Hence, observations of the piezometric head in a set of piezometers can be used to draw isolines of the piezometric head and, subsequently, lines perpendicular to them which are streamlines. Then, by knowing the depth of the saturated zone and using Darcy's law, one can compute the discharge (i.e., the volume of groundwater flow per unit time) between any two streamlines.

Darcy's law has also been derived analytically. Indeed, a fluid moving through the void space of a porous material is subjected to the driving forces of gravity and pressure differences on the one hand, and to the drag force of the solid phase which opposes the movement on the other hand. The piezometric head at any point represents the sum of the gravitational and pressure energy per unit weight of the fluid at that point, and the drop of energy per unit length of a flow path is the sum of the driving forces. Assuming that the drag force is proportional to the specific discharge and the acceleration of the fluid in the neighborhood of a point is negligible, one arrives at Darcy's law.

Darcy's law, wherever applicable, is widely used in groundwater flow models to represent flows across subsurface boundaries within the saturated zone, water exchange between groundwater and surface water bodies (rivers, lakes, the sea, reservoirs) as well as outflow to springs and drainage facilities.

Further extensions of Darcy's law were also made for the cases of an anisotropic aquifer, and a nonhomogeneous fluid. It is worth noting that, in an anisotropic aquifer, the direction of flow and hence also of the specific discharge does not necessarily coincide with the hydraulic gradient, which makes the computations more complicated. In the case of a fluid of variable density, the piezometric head becomes irrelevant and is replaced by the pressure as the relevant state variable.

At this point it is worth becoming familiar with two parameters which are important in formulating a hydraulic, essentially horizontal, groundwater flow model:

- Storativity or storage coefficient: the volume of water an aquifer releases from (or takes into) storage per unit surface area of the aquifer per unit decline (or rise) of the piezometric head. Storativity, at a given point in the horizontal plane, equals the specific storativity, defined earlier, multiplied by the saturated thickness of the aquifer at the given point;
- Transmissivity: the discharge of groundwater through a unit width of the saturated zone as a result of a unit hydraulic gradient. Transmissivity, at a given point in the horizontal plane, equals the hydraulic conductivity, defined earlier, multiplied by the saturated thickness of the aquifer at the given point. In an anisotropic aquifer, the transmissivity at a point depends on both the direction of the normal to the area of the unit width, and the direction of the unit slope. In an isotropic aquifer, the transmissivity is a number independent of direction.

Aquifers and Groundwater as Part of a Water Resources System

At early stages of groundwater exploitation, withdrawals and uses of water in a region are sparse, and their rates constitute a negligible fraction of the total discharge. Under such conditions the natural components of the mass balances of groundwater and its constituents in the region remain essentially undisturbed. However, as populations grow and concentrate on small tracts of land, and as economic development advances, withdrawals and uses of water increase in intensity and variety. As a rule, human settlement and economic activity are also accompanied by uses and releases of substances that, either directly or indirectly, reach natural water bodies, including groundwater. Thereby a new type of a complex water resources system is formed in which natural and anthropogenic components interact with each other and modify the original state of the natural water sources. A schematic diagram of such a system in a region is depicted in Figure 3.

The links of the aquifer-groundwater subsystem with other parts of the water resources system can be used in satisfying human needs related to water. Thus, as stated above, groundwater may serve as a reliable and durable source of fresh water for various uses; groundwater levels can be manipulated to regulate springflow as well as the exchange of water and solutes between aquifers and adjacent surface water bodies (streams, lakes, sea). Costly water conveyance and distribution systems can be avoided by utilizing aquifers as conduits for transporting water over large distances and/or to individual users spread out over large areas.

Contrary to the favorable effects listed above, interactions between groundwater in a region and its environment may also produce effects unfavorable to both the groundwater and the environment. The most common of these effects are:

1. Unfavorable effects of the environment on groundwater:
    - Pollution of groundwater by indirect recharge from anthropogenic sources of pollution via the unsaturated zone and/or by release of pollutants that have been accumulated in that zone;
    - Contamination of groundwater by influx of contaminants from natural surface or subsurface water bodies;
    - Contamination of groundwater by direct injection of contaminated water;
    - Accumulation of contaminants in areas of groundwater level depressions caused by concentrated pumpage;
    - Loss of groundwater to undesired outlets, e.g., contaminated surface water bodies or aquifers;
    - Continued decline of groundwater levels by reduction of recharge as a result of land-use alterations and/or by excessive withdrawal;

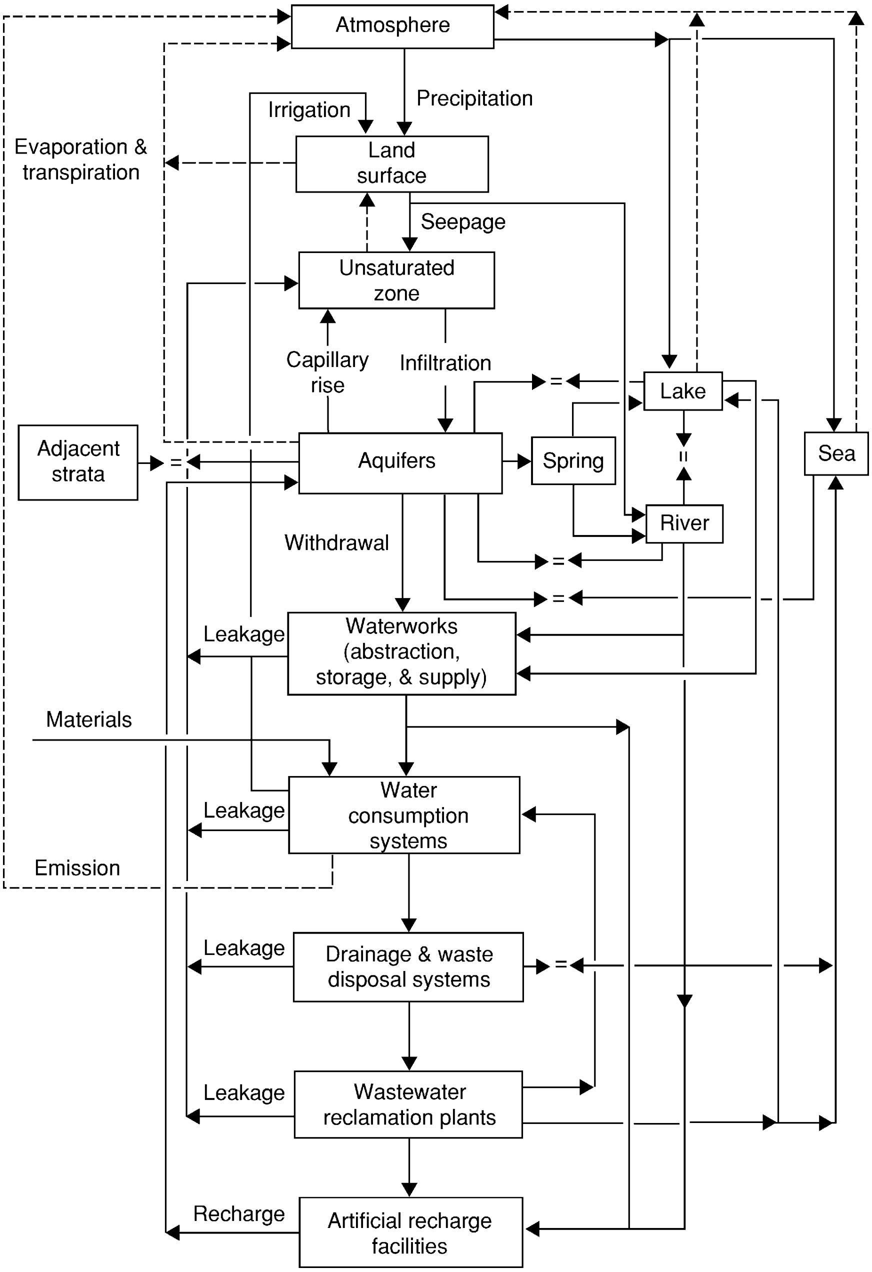


Figure 3 Water resources system. (Adapted from Bachmat Y (1994) Groundwater as part of a water system. In: Zoller U (ed.) *Groundwater Contamination and Control*. New York: Marcel Dekker, Inc. with permission.)

2. Unfavorable effects of groundwater on the environment:
   - Abandonment of pumping wells in areas of contaminated groundwater;
   - Contamination of streams or lakes by contaminated groundwater;
   - Damage to vegetation, crops, soils, structures, and land surface caused by a rise of the groundwater table and/or by deterioration of groundwater quality;
   - Decline of flow and/or water levels in streams and/or lakes as a result of reducing the outflow of groundwater to these water bodies;
   - Reduction or cessation of springflow as a result of declining groundwater levels;
   - Reduction of groundwater outflow to adjacent regions as a result of lowering the groundwater levels in the region;
   - Land subsidence caused by depletion of an underlying compressible confined aquifer.

All of the unfavorable effects listed above may have repercussions on health, material wealth, and other aspects of well-being of the population, not only within the region under consideration, but also beyond it.

Groundwater Contamination Sources and Events

The natural processes and anthropogenic activities that generate groundwater contamination are many and varied, and so are their sources. The sources of groundwater contamination are grouped in two categories: point and nonpoint sources.

Examples of anthropogenic point sources are: open solid-waste dumps; landfills; septic tanks; cesspools; injection wells of inferior-quality water: aboveground and underground storage tanks of hazardous and nonhazardous waste materials; gas stations; poorly constructed production or monitoring wells, which serve as conduits for the vertical transfer of pollutants, either from surface sources to the saturated zone or between geologic layers in the subsurface; and spills from transport facilities. The effect of a point source is, in many cases, local and therefore relatively easy to be monitored and ascribed to the source.

Examples of anthropogenic nonpoint sources are land application of chemicals in agriculture, irrigation with poor-quality waters, percolation of atmospheric pollutants, urban runoff, and patterns of groundwater pumpage and/or groundwater contamination by interaction with surface water bodies. These produce large-scale effects in both time and space, which are often overlapping and difficult to monitor and relate to each source separately.

There is a vast literature of reported groundwater contamination events in all parts of the world. Among the naturally occurring sources of groundwater contamination, whose discharge is created and/or exacerbated by human activity, the best known is seawater intrusion in coastal aquifers. Under undisturbed natural conditions, a state of dynamic equilibrium is maintained whereby flow of fresh groundwater to the sea keeps the landward encroachment of seawater into the aquifer within certain limits. However, as pumpage from the aquifer develops, causing reduction of outflow to the sea and subsequent lowering of the groundwater table (or the piezometrtc surface in a confined aquifer) near the coast, the dynamic balance between the freshwater and seawater is disturbed, permitting seawater to advance inland and contaminate usable parts of the aquifer. As long as the level of pumpage from the aquifer does not exceed its replenishment, a new equilibrium may be reached. Otherwise, seawater intrusion will proceed to contaminate an ever-increasing part of the aquifer. It is worth noting that, although the process of seawater intrusion is slow, it is practically irreversible. Field observations have shown that, once an area has been invaded by seawater, all pumping wells in that area will sooner or later be contaminated and abandoned.

Seawater intrusion has been reported in several parts of the world and is regularly followed up in the Netherlands and in Israel. The best-documented cases in the USA include Long Island, Miami, and many parts of California.

In the USA, contamination of groundwater is usually ascribed to the migration of harmful substances from waste-storage sites, landfills, and farms. According to the Environmental Protection Agency (EPA), industrial wastes are the primary source of groundwater contamination in the USA. Numerous cases of groundwater contamination, particularly from landfills, mine-tailings, and fertilizers, have been reported in Canada. In Israel, the focus of attention is on groundwater contamination by chlorides and nitrates in the coastal aquifer, which is one of the country's major sources of water supply. Originally, when pumpage was concentrated primarily along the coastline, the main source of contamination was considered to be seawater intrusion. As of 2003, when the major well fields are farther inland and concentrations of contaminants in pumped water are nevertheless increasing, the major items of attention are anthropogenic sources on the surface of the ground. An investigation of groundwater contamination by nitrates in a typical part of the coastal aquifer, based on data from 391 wells over an area of $87\,\text{km}^2$, has shown that approximately 61% of the nitrogen in the groundwater originates from fertilizers, 16% from irrigation water, 10% from sewage and refuse, and the remaining 13% is shared almost equally by groundwater inflow through lateral boundaries, urban runoff, and rainfall.

In several European countries with intensive farming (France, Hungary, Russia, the UK, Czechoslovakia, Germany, the Netherlands), a close statistical relationship has been found between the amount of nitrogen fertilizer applied to farmland and the nitrate concentration in groundwater, indicating that under current fertilization practices nitrogen input exceeds its uptake. Contamination of groundwater from domestic, municipal, human, and animal waste is a major problem in many countries in Africa, Asia, and Latin America. Parts of urban and rural environments in many developing countries, such as India, Kenya, Nigeria, Sudan, Egypt, Iraq, and Brazil, are being contaminated by waste products generated in large quantities as a result of rapid industrial development. Intensive, uncontrolled use of fertilizers and poor farming practices is also reported to be a major cause of groundwater contamination in developed countries.

It is worth noting that, except for particular cases of direct contact between local sources and outlets of water and/or contaminants, the occurrence of any of the above effects is not immediate, namely, there is a time lag between the onset of the cause and the occurrence of its effect. This fact is of particular significance with regard to the contamination of groundwater and/or its impact on the environment. Thus, a pollutant released from a surface source of contamination, within a groundwater recharge area, may reach a certain pumping well after many years, or it may never be observed there. Indeed, the path of such a pollutant and the evolution of its concentration are determined by a multitude of factors, such as the depth of the unsaturated zone, the flow pattern within the saturated zone and, hence, the distance to be traveled by the pollutant within this zone toward the pumping well, the pattern of pumpage (locations and rates), and, finally, the properties of the two zones that affect the rate of transport and transformation of the pollutant.

It may thus be concluded that processes of groundwater contamination may remain undetected for long periods of time, turning short-term benefits from invisible polluting activities, at present or in the past, into visible long-term losses from these activities in the future. Similarly, once contamination of groundwater and/or its impact on the environment has been detected, it may require a very long and costly effort to see the desired result of remedial actions if it is at all practically realizable.

The case of the Aswan Dam in Egypt serves as an example of groundwater pollution problems and environmental damage that can be exacerbated by man-made modifications of the natural water system. Groundwater levels in lower Egypt rose considerably with the building of the Aswan Dam in 1965. The regional rise in water levels of shallow aquifers compounded the problems of pollution spread through subsurface disposal of wastewaters and irrigation water. It was reported that "villages are experiencing considerable difficulties with wastewater disposal due to subsurface saturation, high groundwater, emerging surface pools of septic waters, gross groundwater pollution, deterioration of buildings and structures due to moisture absorption and other related problems."

Groundwater and Aquifers Need to be Managed

The benefits that can be drawn from the links of groundwater and aquifers with their physical environment, on the one hand, and the possible adverse effects of those links, on the other hand, lead to the conclusion that the behavior of groundwater and aquifers needs to be controlled. In order to give it a chance, groundwater within a given administrative region has to be dealt with concurrently as both part of a basin and part of the water resources system in the region. Assuming that the overall objective is to improve the economic well-being and the environmental conditions of the region's population, the task of the water resources system, including its groundwater subsystem within that region, would be to supply water for direct consumption (e.g., drinking, recreation), for the production of (industrial and agricultural) goods and power, and for the provision of services (e.g., navigation). However, the latter two are inevitably accompanied by application and release of contaminants which, as shown earlier, may contaminate and lower the intended economic value of the water and/or harm the health of direct users of the water. Uncontrolled abstraction of groundwater as a source of water supply may also be accompanied by losses from mining groundwater reserves and from influx of contaminants from other natural or anthropogenic sources.

This leads to a second conclusion, namely, that an appropriate balance has to be found between actions aimed at improving the economic well-being of the population in a region and those aimed at securing the quality of the environment, including the natural waters. The term 'appropriate' is used here in the sense that all parties and interests involved are satisfied to a degree that no one's well-being can be further improved without hurting someone else. An adequate framework for accomplishing this task may be an integrated regional water resources management system, whose function is to plan and implement means of control, related directly or indirectly to water, in a way that renders the highest net benefit from the water resources system to the entire population of the region. The activities concerned include withdrawal of water from aquifers, lakes, and streams; artificial recharge of the above; import and export of water; water treatment; reclamation and reuse of wastewater; allocation and mode of supply of water from the various sources to uses and users; land-use activities (e.g., placement, construction, operation, and maintenance of facilities and land application of substances which may contaminate the water sources); and transfer of water between components within the water system of the region, and between them and the environment.

As groundwater is part of the water resources system in a region, the management of this system should also comprise a groundwater management subsystem. Its particular task should be to plan and implement the control of groundwater quantity and quality so as to maximize the excess of gains from meeting groundwater-related demands, both present and future, over losses incurred by adverse effects of links with the environment within the region and across its boundaries.

Groundwater Management Levels and Models

The groundwater management process is accomplished via a three-level system: (1) the strategic level, which sets long-range objectives, determines decision criteria and constraints, and prescribes policy guidelines; (2) the tactical level, which translates the directives of the strategic level into long- and short-term plans and projects; and (3) the executive or field level, which controls the specification and implementation of the projects.

Each of the above levels of the management process involves the employment of three categories of formal models. First among them are dynamic models, which are used to predict the behavior of the groundwater system in terms of relevant state variables of groundwater quantity and quality under a given pattern of groundwater operation and/or development.

The starting point of building a dynamic model is a hydrogeological investigation aimed at locating the aquifer and the groundwater stored in it, as well as identifying and locating sources and outlets through which water and its constituents can be exchanged between the aquifer and its physical environment. This is followed by performing a sequence of tasks which is summarized in the form of a list of statements and assumptions that constitute the conceptual model of the system. The main items comprising the conceptual model are:

1. Boundaries, subdivision, and interconnection of aquifers and aquitards in the region of interest;
2. Substances of interest (e.g., groundwater, solutes, and/or suspended matter which are pertinent to water quality requirements);
3. State variables which are relevant to the substances (e.g., groundwater density, piezometric head, temperature, concentration of solutes);
4. Sources and outlets of the substances within the domain of interest;
5. Processes of transport and transformation that are relevant to the substances (e.g., groundwater flow, evaporation, mass transport of solutes, chemical reactions within the liquid phase, mass transfer between the solid matrix and the liquid phase) within the domain of interest;
6. The environment of the considered domain and the processes of exchange of mass of the substances and/or heat across the boundaries with the environment;
7. Assumptions about relevant physical, chemical, biological and other properties of the substances.

The next step in the process of building the dynamic model is the translation of the conceptual model into a mathematical statement referred to as the mathematical model of the groundwater-aquifer system. The main items comprising the mathematical model are:

1. Mathematical formulation of the configuration of the boundaries of the system;
2. Formulation of the balance equations of mass (and heat wherever needed) of the substances of interest;
3. Formulation of the transport equation for each substance in terms of relevant state variables;
4. Formulation of source and sink functions of each substance of interest in terms of observable quantities and/or state variables.

   Upon inserting (3) and (4) into (2), the balance equations take on the form of a set of so-called field equations of the state variables of interest:

5. Formulation of initial values of the relevant state variables;
6. Formulation of the boundary conditions which express the transport of each substance across the boundaries of the system in terms of relevant state variables.

The solution of the field equations under the given initial and boundary conditions takes the form of fields of the state variables of interest within prescribed time-and-space domains of the problem at hand. Consider, for example, the construction of a dynamic model of groundwater flow in a given aquifer. The conceptual model of the system states that the substance of interest is groundwater and the variable of interest is the piezometric head. The lateral boundaries of the aquifer form a rectangle; at one side of the rectangle the aquifer is in contact with the sea, whereas all other sides are impervious. The upper boundary of the saturated zone is the water table; the base of the aquifer is horizontal and impervious. The aquifer is

Table 1 Sources and outlets of groundwater in the selected region

Inflows	*Outflows*
Replenishment from rainfall	Pumpage
Return flow from irrigation	Outflow to the sea
Leakage from water supply system on the surface of the ground	
Inflow through artificial recharge wells	

homogeneous, isotropic, and rigid. The groundwater is incompressible, and the density and temperature of the groundwater are uniform and constant.

The sources and outlets of the groundwater are presented in Table 1. The flow of the groundwater is essentially horizontal. Therefore, and owing to the homogeneity of the aquifer and the groundwater, it is sufficient to describe the behavior of the groundwater at the hydraulic level.

The formulation of the mathematical model starts with the volume balance equation of the groundwater (since the density of the groundwater is constant). Table 1 lists the components of the water balance which change the volume of groundwater in the aquifer. The term 'groundwater storage' is often used to denote the total volume of groundwater in the region of interest. The verbal formulation of the global groundwater volume balance reads:

$$\begin{aligned}&\text{Rate of increment of groundwater storage}\\&\qquad= \text{Sum of the rates of all inflows}\\&\qquad\quad-\text{Sum of the rates of all outflows}\end{aligned}$$

Table 1 shows that only two components of the volume balance are fully controllable: pumpage from wells, and artificial recharge through wells. The rest are not controllable, either fully (replenishment from precipitation) or partly (return flow from irrigation, leakage from surface sources, outflow to the sea). Hence, in order to make the balance equation usable for the purpose of predicting the anticipated change in the state of groundwater (i.e., groundwater level, groundwater flow rates and directions) as a result of pumping or recharge operations, one has to express the uncontrollable terms as functions of relevant measurable quantities (e.g., natural replenishment as a linear function of precipitation depth, return flow from irrigation as a linear function of irrigation depth) and/or relevant state variables (i.e., outflow to the sea and change in groundwater storage as functions of groundwater levels). This step requires utilization of existing scientific knowledge and adaptation of it to the specific conditions in the area of interest. In this example it implies using Darcy's law with the coefficient of transmissivity and the change in storage with the coefficient of storativity defined earlier.

The last step in building the dynamic model is fitting it to the aquifer of interest. This step, referred to as calibration of the model, includes the selection of model parameter values (e.g., the values of transmissivity, storativity, coefficient of natural replenishment and coefficient of return flow from irrigation, in the present example) which ensure a best fit between the historical data of groundwater levels and concentrations with data predicted by the model.

Prediction of future changes of groundwater storage and quality requires data about the configuration, composition, and present state of the groundwater system, data on future inputs of water and contaminants from surficial sources, and also quantitative knowledge about the processes of transport and transformation that may occur within the groundwater subsystem and on its boundaries. Acquisition of this information is hampered by several factors, some of which are specific for a groundwater system. Indeed, any groundwater-aquifer system is hidden and, hence, inaccessible to direct observation. This results in uncertainties concerning its configuration, composition, sources, and outlets. Also, the actual state of the system at a given time is not directly observable. It can only be inferred from assumed relationships with observations of relevant quantities made at discrete locations and times. In addition only a few of the external links of the groundwater subsystem are almost fully controllable. These include direct pumpage and/or recharge. All other links can be grouped into three categories: (1) those that are practically uncontrollable, such as precipitation; (2) those that can only be indirectly controlled by groundwater management, such as flow and mass transfer between the groundwater and external water bodies, whether on the surface of the ground or in the subsurface; and (3) those that are primarily controlled by human activities on the surface of the ground (e.g., irrigation, waste disposal, land-use alterations).

Consequently forecasts of dynamic models are never exact and can at best be expressed in probabilistic terms only. But the error of prediction can still be statistically reduced as more relevant observations are gathered and utilized in the various stages of the modeling process.

The second kind of models are evaluation models, which assign to any course of action and any possible state of the system a set of values of preselected quantitative measures of success ('attributes' or 'criteria') in pursuing the objective(s) of managing the system. The evaluation model can thus be utilized to evaluate the outcome of each alternative course of action, as predicted by the dynamic model. It should be noted that the evaluation model is also not perfect, for reasons analogous to those of the dynamic model. However, like a dynamic model, its level of reliability can also be raised by means of relevant observations.

The third type of model is the model of choice. It receives the output of the evaluation model for different courses of action and selects according to a chosen method the alternative that is most attractive from the point of view of reaching the objectives of the system. The model of choice may have to cope with two problems: uncertainty in the information received from the evaluation model and multiplicity of noncommensurable attributes such as measures of health, level of contamination by various contaminants, and financial loss.

All three of the above models are sometimes included in a single optimization model. Although formally attractive, the use of optimization models in practice is limited. The reasons for this situation are twofold: The limited capability of an optimization model to represent a complex real system, and the reluctance of decision makers to accept a prescribed course of action without taking part in the process of making a choice.

At the strategic level, one model may represent an entire water sources system. The groundwater subsystem may be represented in this model by several basins, each as a single cell. The same may hold for the surface-water subsystem. Another model at the strategic level may represent the water-supply system, a third, the wastewater disposal and reuse system, and a fourth, the economics of the water resources system. All of the above models are linked to each other. The focus of interest is not so much on the details of each of the subsystems but rather on their interrelation and the resulting long-term consequences.

At the tactical level, each subsystem is modeled separately with an emphasis on the behavior of each of its parts and their interrelations. Obviously, models representing the same system at different levels of aggregation must satisfy certain conditions of compatibility. Of particular interest, in this regard, may be the actual application of a set of such models in Israel in developing a strategy of integrated management of its water resources using the systems approach and multilevel modeling.

To be applicable and useful in practice, all of the above models need data for their formulation, calibration, validation, and operation. This, in turn, requires data-management models, including models for the design of observation programs, models for data quality control, and models for transforming observations to processed data, which serve as inputs to the aforementioned models.

Use and Usefulness of Formal Models in Groundwater Management

Although much progress has been made in the field of development of the above types of models, their actual use and usefulness in the management of groundwater is still wanting. The main reasons for this seem to be:

1. Lack of a mandatory, clearly defined, and duly implemented methodology and procedures of utilizing quantitative models, and information derived therefrom, in the decision-making process;
2. Low credibility of the models, especially those pertaining to groundwater quality.

The situation can be improved by advancement of knowledge about processes of groundwater contamination and their environmental effects, as well as by adoption and improvement of methods of system identification, modeling, analysis, and monitoring. Progress in the above fields and assimilation of the knowledge in practice require an interdisciplinary effort and present a challenge.

Another hindrance to the effective use of formal models in groundwater management may be relatively loose links between modelers and technical staff of model users, as well as between the latter and the managers. The remedy for this is practical training in the fields of interdisciplinary communication and interactive formulation and solution of problems.

Anthropogenic Constraints on Groundwater Management

No management system operates in a vacuum. The same applies to the management of water resources in general and groundwater in particular. Laws and regulations with respect to water, rules prescribed by the socioeconomic and political system of government, along with limited resources and know-how, are constraints that restrict the feasible action space of the water management system.

Legal constraints are often a major obstacle to efficient management of groundwater. For example, despite the fact that groundwater and surface water are interrelated, several states in the USA still legally define the two as being physically disconnected. But the most powerful legal constraints are the property rights on sources of water. Improperly defined property-right systems preclude efficient utilization of groundwater and stimulate overexploitation and contamination.

One type of improperly defined property rights, which, for example, still exists in certain states in the USA, treats groundwater as a common property resource, i.e., as a resource that is not exclusively controlled by a single authority, but by many individual owners, each of whom is entitled to full control over drilling and extraction of groundwater within the limits of a surface area, which is small in comparison with the geographic extent of the relevant aquifer. Although the actions of individual users affect the state of the groundwater beyond their area, the property right protects them from bearing all of the consequences of their actions. Losses from overexploitation and/or pollution of the aquifer, which are taken into account in the case of exclusive ownership, are an externality to the individual owner, and not a part of their decision-making process. Hence, wherever property rights treat groundwater as a common property resource, it cannot be managed efficiently, because the property rights allow unlimited use of the resource and discourage conservation.

Another type of property right that precludes efficient management of groundwater treats water as a public good, which is fully accessible to all, as if it were an unlimited resource such as air. It is claimed that legal constraints of the aforementioned types are being gradually relaxed as the fundamentals of hydrology become better understood by public officials, legislators, and the courts.

The Water Law in Israel has declared all waters as public property subject to state control and intended to be used for the benefit of its inhabitants and for the development of the country. Although, as in this case, exclusive property rights on water are a necessary condition for efficient management of water resources, they are not yet sufficient. Political and administrative constraints, such as the placement of the water management system within the system of government, and the role of the central government in the management process, may have a major effect on the effectiveness of control over this vital resource.

In Portugal, for example, management of water affairs is under the Secretariat of State for Environment and Natural Resources, which resides within the ministry of Regional Development and Land Use Planning. The placement of water management within an intersectorial ministry, which must integrate different strategies of different sectors in the economy, eliminates the possibility of one sector's interests prevailing and resulting in an allocation of water that is inefficient from the point of view of the public at large. Portugal is also restructuring the actual administration of water affairs. A multilevel, decentralized management system is being introduced, in which a national authority (National Water Institute) representing the government sets the rules and procedures of water management, but the actual management of the system rests with the users through the delegation of power and responsibility to regional and local authorities.

France is another example of a country in which management of water resources is decentralized and practiced on the scale of a river basin. The agency that manages the basin, in terms of water use and water quality control, acts on behalf and under the supervision of the users in the basin, but subject to rules set by the government.

Standards of groundwater quality are often imposed as mandatory constraints. Unlike standards on water quality at the site of use, which are imposed by the respective branches of government and rarely questioned by the public, ambient water quality standards in the aquifer are closely related to the intended use of the water and to its treatment after extraction from the aquifer. But decisions of this kind should be derived on the basis of an analysis of options, outcomes, and consequences. Otherwise, arbitrarily imposed ambient standards may lead to an inefficient allocation of the groundwater resource and heavy financial losses.

Nonphysical Means of Groundwater Management

Apart from physical means of control that can be applied directly to the groundwater subsystem (pumpage and/or recharge) and/or to its physical environment (e.g., technological devices and facilities), there is a variety of nonphysical means of control that can enhance the effectiveness of groundwater management. These may include any combination of the following categories:

- Managerial (e.g., enhancement of public participation in the processes and procedures of decision-making, follow-up, and control of implementation);
- Economic (e.g., prices, charges, taxes, rebates, loans, subsidies);
- Legal (e.g., modification of laws and regulations);
- Administrative (e.g., allotments);
- Political (e.g., treaties);
- Educational (e.g., dissemination of information, provisions of guidance, training);
- Scientific (advancement of knowledge about processes, technologies, and techniques).

Experience shows that economic market measures of control are more flexible in responding and adjusting to changes in the quality of the receiving water body and in the benefits that can be derived from pollution abatement than administrative measures. A switch from regulatory measures of water-pollution control to economic incentives, or an appropriate mix of the two, may therefore be desirable. In the USA, for example, a number of empirical cost-effectiveness studies of water-pollution control have shown that the current EPA uniform effluent standards are much less cost-effective than emission charge or marketable emission-permit systems. Evidently, economic incentive systems have a distinct advantage over bureaucratically defined standards by forcing the sources of pollution continually to search for new opportunities to control emissions at lower cost. In Europe, for example, effluent charges play a prominent role in a number of countries.

It should, however, be emphasized again and again that any set of available means of control can serve as effective management tools if, and only if, their choice and implementation are guided by a rational decision-making process that utilizes information on the present state of the system, on the anticipated effect of any undertaken action on the states of the groundwater subsystem and its environment, as well as on the ensuing benefits and losses, and only then selects that combination of means and those levels of actions that are most effective in reaching the groundwater-management objectives. In doing so, losses associated with uncertainties concerning the future state of the system, and the implementation and impact of the undertaken measures of control, should be taken into account. Since these losses can be reduced as information on the state of the system is being updated, it follows that the management of groundwater, as part of a modern water system, should be based on adaptive control by feedback with information from ongoing monitoring of its actual state.

Finally, even the most rational decisions are useless if not duly implemented. Problems of coordination, adherence to schedules, methods, and procedures, access to data, technology transfer, and delays in allocation of funds are commonplace. Therefore, the physical and managerial parts of the groundwater subsystem should be managed together, maintaining a permanent control over the adoption and actual implementation of decisions by means of an integrated multidisciplinary, multipurpose information service.

Acknowledgment

This article is adapted with permission from Bachmat Y (1994) Groundwater as part of the water system. In: Zoller U (ed.) *Groundwater Contamination and Control*. New York: Marcel Dekker.

See also: **Darcy's Law**; **Drainage, Surface and Subsurface**; **Hydrodynamics in Soils**; **Pollution:** Groundwater; **Vadose Zone:** Hydrologic Processes

Further Reading

Bachmat Y (1989) *Management of Groundwater Observation Programs*. World Meteorological Organization, Operational Hydrology Report No. 31, WMO-No. 705. Geneva, Switzerland: WMO.

Bachmat Y (1994) Groundwater as part of the water system. In: Zoller U (ed.) *Groundwater Contamination and Control*. New York: Marcel Dekker.

Bachmat Y and Chetboun G (1974) *Seawater Encroachment in the Coastal Plain of Israel*. Jerusalem, Israel: Israel Hydrological Service.

Bear J (1979) *Hydraulics of Groundwater*. New York: McGraw-Hill.

Bear J and Bachmat Y (1991) *Introduction to Modeling of Transport Phenomena in Porous Media*. Dordrecht, the Netherlands: Kluwer Academic.

Council on Environmental Quality (1980) *Environmental Quality – 1980*. Washington, DC: Government Printing Office.

Domenico PA and Schwartz FW (1990) *Physical and Chemical Hydrogeology*. Chichester, UK: John Wiley.

Environment Canada (1990) *Water Fact Sheet 4*. Ottawa, Canada: Inland Waters Directorate.

Haimes YY (1977) *Hierarchical Analyses of Water Resources Systems*. New York: McGraw-Hill.

Hall WA and Dracup J (1970) *Water Resources Systems Engineering*. New York: McGraw-Hill.

Mandel S and Shiftan ZL (1981) *Groundwater Resources Investigation and Development*. New York: Academic Press.

Mercado A (1980) Israel as a case study. In: Shural HI (ed.) *Water Quality Management Under Conditions of Scarcity*, pp. 93–146. New York: Academic Press.

Office of Technology Assessment (1984) *Protecting the Nation's Groundwater from Contamination*. OTA-0-233 Washington, DC: Government Printing Office.

Piver WT (ed.) (1989) *EHP Monograph on Groundwater Quality*. Washington, DC: US Department of Health and Human Services.

Preul HS (1985) In: *Proceedings of the 5th World Congress on Water Resources*, pp. 357–366. Belgium: Brussels.

Schwarz J (1980) Israel as a Case Study. In: Shuval HI (ed.) *Water Quality Management Under Conditions of Scarcity*, pp. 51–91. New York: Academic Press.

Stokey E and Zeckhauzer R (1978) *A Primer for Policy Analysis*. New York: WW Norton.

Tahal and Mekorot (1987) *Dan Region Project. Treatment and Groundwater Recharge of Municipal Waste-water, 1986 Report*. Tel Aviv: Tahal.

Tietenberg T (1988) *Environmental and Natural Resource Economics*, 2nd edn. Glenview, IL: Scott, Foresman.

UNESCO World Meteorological Organization (1992) *International Glossary of Hydrology*, 2nd edn. Paris, France: UNESCO, WHO.

van der Heijde P, Bachmat Y, Bredehoeft J, Andrews B, Holtz D, and Sebastian S (1985) *Ground-water Management: The Use of Numerical Models*. American Geophysical Union, Water Resources Monograph 5. Washington, DC: AGU Publications.

Vardi Y (1980) Israel as a case study. In: Shuval HI (ed.) *Water Quality Management Under Conditions of Scarcity*, pp. 37–49. New York: Academic Press.

Willis R and Yeh WG (1987) *Groundwater Systems Planning and Management*. Engelwood Cliffs, NJ: Prentice-Hall.

Groundwater Pollution *See* **Pollution:** Groundwater

H

HEAT AND MOISTURE TRANSPORT

R Horton, Iowa State University, Ames, IA, USA
A Globus, Agrophysical Research Institute,
St Petersburg, Russia

In temperate climates, prompt germination, rapid emergence, and vigorous early growth of crops are critical to achieving profitable yields. Management strategies to deal with cool, wet soil conditions include the timing of tillage and planting operations and decisions regarding residue management, fertilization, and pesticide application. Heat and moisture transfer helps to define the conditions under which each of these management strategies may be most effective. Recent studies have highlighted the effects of temperature and wetting-and-drying cycles on soil C and N dynamics. Microbial dynamics affecting C and N cycling, including the generation of greenhouse gases, are influenced by drying–rewetting cycles and through a temperature–soil-water content interaction. Knowledge of heat and moisture transfer in soil is useful for predicting the soil-water content and temperature conditions that influence microbial dynamics in the C and N cycles. Understanding coupled heat and water flow processes, therefore, will enable significant and wide-ranging improvement of management strategies to optimize production efficiencies and reduce environmental impacts.

Coupled transfer of heat and water in soil is common. By coupled transfer we mean that heat and water transfer processes are intertwined. When water moves in soil, it carries heat with it, and temperature gradients that develop in response to heat transfer cause both liquid water and water vapor to move. Diurnal and seasonal climatic fluctuations cause coupled heat and water transfer to occur. For example, in a typical diurnal cycle in an agricultural field, the soil loses heat during the night. As heat leaves the soil, the surface temperature drops, and an upward thermal gradient develops, resulting in transfer of heat to the surface from the relatively warm soil below. If the surface temperature drops to the dewpoint temperature, water vapor condenses at the cooled soil surface, thus liberating heat. Throughout the night, liquid water flows upward toward the surface, rewetting the surface layers that dried during the previous day and redistributing heat upward. In the morning, solar radiation warms the soil surface. As heat enters the soil, the surface temperature increases, and the thermal gradient reverses direction. Liquid water near the soil surface vaporizes and is transported into the atmosphere, taking large quantities of heat with it. Liquid water and some vapor are also driven by the thermal gradient from the hot surface soil to the cooler soil below. As the sun goes down, this dynamic and intricate cycle begins again. All soil biological, chemical, and physical processes are influenced by the fluctuations of soil-water content and soil temperature that result from coupled heat and water transfer processes.

Although it has been known for many years that coupled heat and water transfer occurs in soil and that the coupled processes have important consequences for near-surface physical, chemical, and biological activities, the coupled processes have not been well quantified. Improved understanding of coupled heat and water transfer is needed to provide the basis for more complete understanding of soil-water and temperature conditions, soil-water evaporation, crop and weed seed germination, pesticide volatilization, surface fluxes of carbon dioxide, and trace gas emissions from soil. Improved understanding of coupled heat and water transfer is also needed to provide the basis for further understanding of the impact of climate change on the near-surface soil environment.

Temperature gradients exist in soil owing to periodicity of insolation, geothermal temperature distribution, functioning of buried cables, heating and cooling pipes, etc. Existence of temperature gradients means violation of thermodynamic equilibrium, which results in fluxes of heat and water in soil. The water flux can induce flux of soluble substances.

Soil moisture transfer induced by temperature gradient is called the thermal moisture transfer (TMT). The flow resulting from interaction of TMT with moisture flow induced by factors other than temperature gradient is called the nonisothermal moisture transfer (NIMT). TMT can be vapor, liquid, and combined (series-parallel).

Vapor TMT is mainly caused by dependence of water vapor concentration (and/or pressure) on temperature, thus gradients of temperature create gradients of water vapor concentration and hence classic thermal water vapor diffusion in soil, which is directed toward a lower temperature. As with normal gas diffusion, vapor TMT depends upon the liquid-free porosity and on tortuosity of diffusion path, which is influenced by soil texture, structure, and bulk density.

Liquid TMT can be effected by several mechanisms:

1. Expansion and contraction of entrapped air due to temperature change pushes liquid back and forth in nearly saturated soil, for example, in the capillary fringe and near the groundwater table. This is the so-called thermometric effect, which is transient;
2. Due to dependence of surface tension at the liquid–air interface on temperature, a temperature gradient induces a gradient of surface tension and respective gradient of capillary pressure of menisci. This can induce hydrodynamic flow in the same way as the influence of a pressure difference of any other origin. This is 'thermocapillary meniscus' flow;
3. When a thermal gradient exists along the liquid–air interface of a liquid film covering solid particle(s), the induced surface tension gradient produces thermocapillary film flow. The velocity profile of this flow (as a function of distance to solid phase), differs from that of hydraulic flow, because the moving force is applied only to the interface. Both thermocapillary flows are directed toward a lower temperature;
4. When enthalpy of pore liquid differs from that of bulk liquid, and, particularly, when there exists some distribution of enthalpy as a function of distance to the solid phase, thermo-osmotic flow can occur. Direction as well as magnitude of this flow depend upon the actual profile of enthalpy. Since normally the latter diminishes near the solid–water interface, thermo-osmotic flow should be directed toward a higher temperature;
5. In unsaturated soil there exists a special vapor–liquid series-parallel (or combined) TMT, consisting of thermal vapor microdiffusion inside air space of a pore, combined with liquid flow in capillary and film elements of soil matrix. Because of its dual nature, this flow bears features appropriate to both vapor diffusion and to liquid flows. For example, it depends upon ambient gas pressure P as vapor diffusion (diminishing with rise of P), and on phobicity of a medium to liquid flow, vanishing in hydrophobic media. Due to the same dual nature of this mechanism, dense and friable soil layers are almost equally well permeable for the combined TMT, supposedly because the increasing role of the faster liquid component in the denser layer compensates for the reduction of water-free space for vapor diffusion. This property of combined TMT can be used for regulating the root-zone water regime, since a friable interlayer, transparent for downward TMT, is less permeable for upward, 'isothermal' liquid flow, thus creating a trap for the moisture transferred by TMT.
6. A special case of moisture migration in soil under the influence of a temperature gradient is the movement of moisture in soils induced by freezing. This process is often described as TMT, where the driving force is formally represented by the temperature gradient. However, actual water flow in this case is caused by the gradient of chemical potential of water, which arises due to locally reduced liquid water content at and behind the freezing front after freezing of some water. Differences in unfrozen water content create differences in matric potentials. So in this case it would be proper to substitute temperature gradient by hydrostatic pressure gradient with the help of the appropriate form of the Clausius–Clapeyron equation. Just like freezing itself, this type of 'TMT' can be suppressed by salinization and/or overburden pressure, which increase the chemical potential of water and reduce the freezing point of pore solution.

Phase composition of TMT governs its contribution to heat transfer, since pure vapor and combined flows participate in transfer of latent heat (approximately 2500 kJ kg^{-1} of water); meanwhile, liquid TMT carries only 4.2 kJ kg^{-1} of water per unit temperature difference. The ratio of vapor and liquid contributions to total TMT depends on the soil moisture content.

There exist three interrelated types of mathematical description of TMT: Fick's formalism, Darcy's formalism, and formalism of irreversible thermodynamics: $q_{wt}/\rho_w = -D_T\nabla T$; $q_{wt}/\rho_w = -K_T\nabla T$; $J_{wt} = L_{WT}X$, respectively, where: q_{wt} is time rate of TMT, ρ_w is density of water, D_T is thermal moisture diffusivity, K_T is thermal hydraulic conductivity, J_{wt} is the same as q_{wt}, L_{WT} is the phenomenological coefficient, X is thermodynamic force (the product $L_{WT}X$ must have dimensions of internal entropy production time rate, so, for example, X can be set equal to $-\Delta T/T^2$, where ΔT is temperature difference on the system boundaries (one-dimensional case), T is mean absolute temperature). Physically, D_T and K_T are the same, though routinely the symbol 'D_T' is in use. D_T can be split into D_{Tv}, thermal vapor moisture diffusivity (usually also accounting for combined TMT), and

D_{Tl}, thermal liquid moisture diffusivity. The same can be done with K_{T} and L_{WT}.

Additionally, $q_{\mathrm{WTf}}/\rho_{\mathrm{w}} = -K_{\mathrm{Tfz}}\nabla T_{\mathrm{f}}$, where q_{wtf} is time rate of water flow in freezing soil, K_{Tfz} is respective 'thermal' hydraulic conductivity, T_{f} is temperature in the range of water–ice phase transition temperatures.

Classic thermal vapor diffusivity $D_{\mathrm{Tcv}} = D_0(\Pi - \theta)\,\gamma\,\varphi\partial c_0/\partial T$, where D_0 is coefficient of water vapor diffusion in air, Π is total porosity, θ is volumetric water content, γ is tortuosity factor (dimensionless), c_0 is concentration of saturated water vapor in air, φ is equilibrium relative humidity of air in soil. The order of magnitude of D_{Tcv} is $10^{-9}\,\mathrm{m^2\,s^{-1}\,K^{-1}}$.

Liquid (thermocapillary) thermal water diffusivity D_{Tcp} is often expressed as $K(\psi/\sigma)\partial\sigma/\partial T$, where K is usual unsaturated hydraulic conductivity, ψ is equivalent hydrostatic head, and σ is surface tension of water on interface with air. This form of D_{Tcp} results from the assumption that the contact angle λ in capillarity law $P = C\sigma\,\cos\lambda/r$ (c is constant, r is radius of equivalent capillary) does not depend on T. If temperature dependence does take place, an additional term will appear in D_{Tcp} in the form of $(\psi/\lambda)\partial\cos\lambda/\partial T$.

D_{TP} for series-parallel transfer can be represented as ξD_{Tcv}, where ξ is enhancement factor, $1 < \xi < 10$, which accounts for increased temperature gradient inside air space of pores and interaction of vapor and liquid flow, which makes the space occupied by the liquid phase 'transparent' for this mechanism of TMT. This enhanced TMT is appropriate not only to water but also to such substances as ethanol, benzene, and carbon tetrachloride. However, the larger the $D_0\partial c_0/\partial T$, the smaller are the respective ratios $D_{\mathrm{TP}}/D_{\mathrm{Tcv}}$ of these substances. All thermal moisture diffusivities increase exponentially with increasing T: vapor diffusivities owing to increases in D_0 and $\partial c_0/\partial T$ with rises in T; and liquid diffusivities owing to reducing viscosity of a liquid with rise of T, which in the range of normal temperatures is relatively greater than diminishing σ. D_{Tcv} and D_{TP} decrease inversely with increase in ambient gas pressure P. Electrolytes in soil solution depress the moisture flux provided by combined TMT. For sodium and calcium chlorides, this effect can be described by the equation

$$(D_{\mathrm{TP}})_{\mathrm{N}}/(D_{\mathrm{TP}})_{\mathrm{N=0}} = 1 - A\sqrt[3]{N}$$

where A is a constant ≈ 0.5 and N is normality of electrolyte solution.

The order of magnitude of K_{Tfz} is $10^{-7}\,\mathrm{m\,s^{-1}}$, so TMT under freezing proceeds much faster than above 0°C, though dehydration of soil next to an ice lens puts an end to the process until the freezing front moves deeper over the freeze-dried zone. Due to moisture transfer toward the freezing front, as well as the freezing of part of the water, the zone before a freezing front accumulates soluble substances.

No matter how large (or small) are the respective values of D_{T} or their sum, within the full range of possible soil moisture contents there exists a range where TMT is comparable with the flux, induced by matric pressure gradient. This range is the domain of effective NIMT, i.e., soil moisture-content diapason where TMT can result in net changes of moisture-content distribution. Experiments with closed systems showed that this range increases with increase in clay content (or specific surface area S), reduction of bulk density, and increase in aggregate percentage. It also depends on temperature-gradient magnitude and on absolute temperature. Inside this range (expressed in terms of saturation degree concept) there often exists a particular 'critical' degree of saturation DS_{c}, at which NIMT has its maximum, depending on the same factors as the range itself. For a temperature gradient of the order $100^{\circ}\mathrm{C\,m^{-1}}$ (often occurring in shallow soil) and ambient temperature, $DS_{\mathrm{c}} = 0.085 + 0.024\sqrt{S}$, where S is in meters squared per gram. In closed or quasiclosed systems, steady-state moisture distribution is achieved owing to a sort of circulating system, when TMT carries moisture to the cold boundary of a system, thus creating a gradient of moisture content and matric potential, causing moisture to flow back to the hot boundary predominantly as a liquid. In summer days a quasiclosed system in terms of NIMT occurs in the upper layer of soil, since temperature gradients, maximal at the surface, gradually vanish with depth. Moisture driven mainly by combined TMT into a cooler layer flows back to the surface as a liquid, carrying with it soluble substances. At night TMT with condensation can partly replenish the moisture supply of the uppermost soil layer. The moisture circulation due to the NIMT can enhance salinization of the uppermost soil layer with accumulation of soluble substances (useful as well as harmful ones) in it.

Examples of Coupled Heat and Water Flow in Soils

Three examples of coupled heat and water flow in soil are presented. One example represents field infiltration of water under nonisothermal conditions, and the other two examples are for closed laboratory soil columns that experience nonisothermal conditions. The field example shows heat and water transfer under wet soil conditions in response to diurnal heating and cooling. The laboratory examples

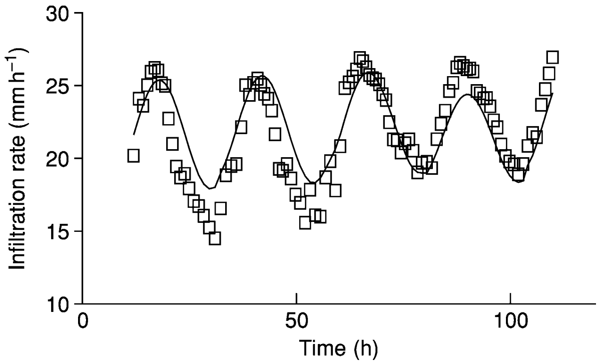


Figure 1 Soil-water infiltration rate as a function of time. Open squares denote measured data, and the solid curve indicates predicted infiltration rates. The model used to predict infiltration rate includes the effect of temperature on water viscosity.

demonstrate the effect of temperature on coupled heat and water transfer and the effect of salinity on coupled heat and water transfer in unsaturated soil.

Figure 1 presents field observations of ponded water infiltration rate as a function of time. The experiment lasted for 5 days. Diurnal fluctuations in water infiltration rate are delineated, and the solid curve represents a simple model of the data. The only parameter included in the model is water viscosity, which is temperature-dependent and thus affects soil hydraulic conductivity. Diurnal soil surface temperature was used to estimate diurnal fluctuations in water viscosity, which in turn was used to estimate diurnal fluctuations in soil-water infiltration rate. As shown in Figure 1, this very simple model captures the trend of the infiltration data and very nearly captures the amplitudes of the diurnal fluctuations. Clearly, heat and water flow occurs in natural environments.

Figure 2 presents laboratory data collected from a series of closed soil columns. Each soil column started at the same uniform initial water content (denoted by the dashed line in Figure 2) and had the same constant thermal gradient imposed on it. The curves in Figure 2 show the water content distributions after a thermal gradient has been applied to each soil column. Although the same thermal gradient was used on all soil columns, the mean temperature varied among the soil columns. For example, mean temperatures were approximately 10°C, 20°C, 30°C, and 40°C. There is a clear effect of mean temperature on the amount of water redistribution in the columns. The mean temperature affects vapor pressure as well as vapor and liquid diffusivities. For large mean temperatures, water transfer can be expected to be large. This has implications for field heat and water transfer as a function of season.

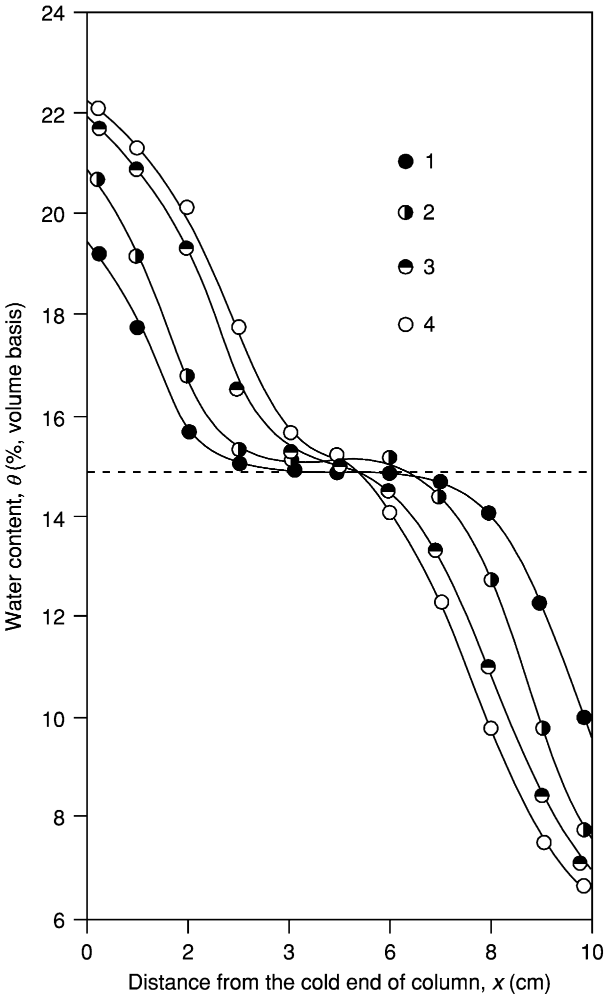


Figure 2 Water-content distributions after soil columns experience the same thermal gradient imposed at different mean temperatures: 1–10°C, 2–20°C, 3–30°C, and 4–40°C. Reproduced with permission from Globus AM (1983) Physics of non-isothermal soil moisture transfer [in Russian]. Leningrad, Russia: Gidrometeoizdat.

Figure 3 presents two water-distribution curves for soil columns that have been exposed to the same thermal gradient. Both soil columns had the same initial water content; however, one soil column contained almost pure water, while the other column had soil solution that was salinized. The curves in Figure 3 indicate that greater water redistribution in response to the thermal gradient occurred in the soil column that was not salinized. In the salinized soil column, much less water redistributed in response to the thermal gradient. These data demonstrate the tempering effect of soil salinity on water movement in response to thermal gradients, which is described by the formula given above. Since most natural soils contain some dissolved salts, it is important to account for soil solution concentrations when estimating coupled heat and water transfer.

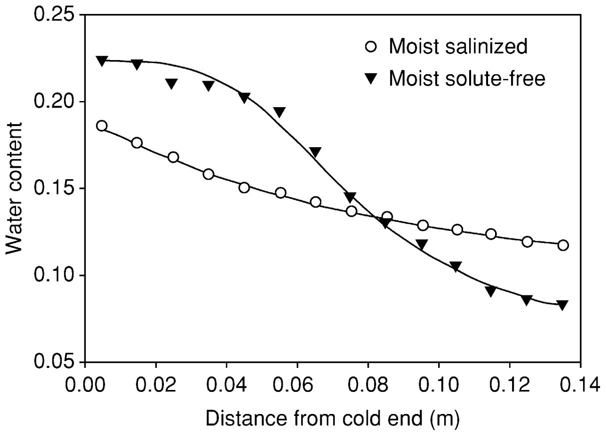


Figure 3 Observed steady-state water-content distributions within moist salinized and solute-free soil columns.

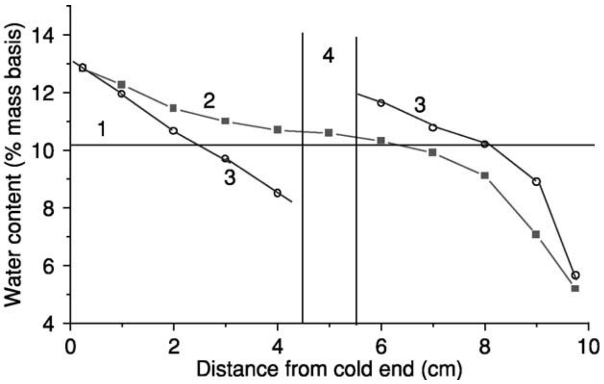


Figure 4 Influence of hydrophobic soil layer on thermal water transfer and moisture distribution in loamy soil. 1, initial moisture content *W*; 2, moisture profile in uniform soil column; 3, the same loamy soil in a column with a hydrophobic layer (4).

Limitations in Our Understanding of Coupled Heat and Water Flow

Although several studies have addressed coupled heat and water transfer in unsaturated soil, fundamental deficiencies in understanding persist. For example, field observations of water vapor movement in soils can be greater than predicted by the widely used diffusion-based theory of Philip and de Vries. Recent measurements highlight our current lack of understanding of the contributions of convection and diffusion to coupled heat and water transfer in soil. Researchers have reported that existing diffusion-based theory is inadequate to describe their field measurements of coupled heat and water transfer. Existing theory cannot explain the large vapor fluxes and evaporation rates measured in some field studies. As a result of these findings, it has been proposed that diurnal heating of the soil causes convective transport of soil air due to temperature-driven expansion of the soil air (which is akin to the thermometric effect, mentioned above). It has also been proposed that this convective transport (soil 'breathing') may explain the large fluxes of water vapor observed in field studies.

Other recent measurements highlight the current lack of understanding of the influence of soil hydrophobicity on coupled heat and water transfer. Current theory of coupled heat and water transfer assumes that soil is perfectly hydrophilic or 'wettable' (i.e., that the contact angle of soil water is 0°). However, recent research suggests that under certain conditions all soils may display water repellency to some degree. Water repellency has been observed in sand, loam, clay, peat, and volcanic ash soils all over the world. Researchers have measured isothermal and nonisothermal (columns heated from below) evaporation from columns containing wettable soil and columns containing hydrophobic soil. Evaporation from the hydrophobic soil was approximately 75% and 50% of the evaporation from the wettable soil under isothermal and nonisothermal conditions, respectively. The available published data of soil wettability effects on coupled heat and water transfer in field soil are limited; however, in laboratory experiments it has been demonstrated (Figure 4) that a porous hydrophobic layer reduces TMT drastically in wet soil. Figure 4 compares NIMT in a uniform column of loamy soil with NIMT in a column of the same soil with a hydrophobic soil layer in the middle section. The data show clearly that soil wettability can have significant impacts on heat and mass transfer in soil. No existing models can describe explicitly these results, although the Philip–de Vries theory assumes implicitly that in hydrophobic soil the series-parallel mechanism does not work.

There are limitations in the models used to predict coupled heat and water transfer in soil. Current models all assume that the state variables can be represented by continuous functions and that volume-averaged transport parameters adequately represent the properties of the soil volume under consideration. These continuum models with volume-averaged parameters have significant limitations for describing coupled heat and water transfer. For example, the widely used de Vries model for soil thermal conductivity requires assumptions about the pore-scale arrangement of soil solids, water, and air. These assumptions are represented in the model by 'shape factors.' The choice of shape factors determines how thermal conductivity varies with soil water content. The lack of a model to describe the pore-scale arrangement of the solids, water, and gas forces us to rely on these arbitrarily chosen shape factors. Hence our ability to predict thermal conductivity is limited, as is our ability to predict coupled heat and water transfer, which is strongly dependent on thermal conductivity. As a second example, Philip and de Vries have made an effort to improve the limitation of their continuum model by introducing a thermal vapor enhancement factor to account for pore-scale features (e.g., nonequal thermal gradients in the various phases and nongas-phase diffusive pathways in the pores). Researchers have used Philip and de Vries' theory to describe nonisothermal evaporation from columns containing wettable soil and columns containing hydrophobic soil. In the most wettable soil column(s), the thermal vapor enhancement factor is approximately 3.5 (meaning that thermal vapor transport is approximately 3.5 times greater than predicted by classic thermal vapor diffusion theory), but in the most hydrophobic soil column the thermal vapor enhancement factor was approximately 0.35.

In other words, soil columns that differ only in wettability show very different water evaporation rates and water redistribution. No existing model can describe explicitly the pore-scale features that lead to these greatly different thermal vapor enhancement factors. The development of discrete models such as pore-scale network models will aid in the understanding and description of coupled heat and water transfer in soil.

Coupled heat and water transfer is very important in soils; it is also very complex. Clearly, it is important for scientists to continue to seek to understand these important, complex processes.

List of Technical Nomenclature

γ	Tortuosity factor
θ	Volumetric moisture content
λ	Contact angle
ξ	Enhancement coefficient
Π	Total porosity
ρ_w	Specific density of water
σ	Surface tension of liquid on the interface with air
ϕ	Relative humidity
ψ	Equivalent (matric) hydraulic head
A, C	Constants
c_0	Concentration of saturated vapor
D_0	Vapor diffusivity in free air
D_T	Thermal moisture diffusivity
D_{Tcp}	Thermocapillary liquid moisture diffusivity
D_{Tcv}	Classical thermal vapor moisture diffusivity for soil
D_{Tl}	Thermal liquid moisture diffusivity
D_{TP}	Series-parallel liquid-vapor thermal moisture diffusivity
D_{Tv}	Thermal vapor moisture diffusivity
DS_c	Critical degree of saturation
J_{WT}	Thermal moisture flux
K	Unsaturated hydraulic conductivity
K_T	Thermal hydraulic conductivity
K_{Tfz}	Thermal hydraulic conductivity due to freezing
L_{WT}	Phenomenological coefficient
N	Normality of electrolyte solution in soil pores
P	Ambient air pressure
q_{WT}	Time rate of thermal moisture transfer
q_{WTf}	Time rate of thermal moisture transfer to freezing front
r	Radius of equivalent capillary
S	Specific surface area
T	Absolute temperature
T_{fz}	Absolute temperature in the range of water–ice phase transition in soil
X	Thermodynamic force

Further Reading

Cary JW (1963) Onsager's relation and the nonisothermal diffusion of water vapor. *Journal of Physical Chemistry* 67: 126–129.

Cary JW (1965) Water flux in moist soil: thermal versus suction gradients. *Soil Science* 100: 168–175.

Cary JW (1966) Soil moisture transport due to thermal gradients. Practical aspects. *Soil Science Society of America Proceedings* 30(4): 428–433.

Cass A, Campbell GS, and Jones TL (1984) Enhancement of thermal water vapor diffusion in soil. *Soil Science Society of America Journal* 48(1): 25–32.

Globus AM (1983) Physics of non-isothermal soil moisture transfer [in Russian]. Leningrad, Russia: Gidrometeoizdat.

Milly PCD (1980) A simulation analysis of thermal effect on evaporation from soil. *Water Resources Research* 7: 1216–1225.

Nassar IN and Horton R (1992) Simultaneous transfer of heat, water, and solute in porous media. I. Theoretical development. *Soil Science Society of America Journal* 56: 1350–1356.

Nassar IN and Horton R (1997) Heat, water, and solute transfer in unsaturated porous media. I. Theory development and transport coefficient evaluation. *Transport in Porous Media* 27: 17–38.

Nassar IN, Horton R, and Globus AM (1992) Simultaneous transfer of heat, water, and solute in porous media. II. Experiment and analysis. *Soil Science Society of America Journal* 56: 1357–1365.

Philip JR and de Vries DA (1957) Moisture movement in porous materials under temperature gradients. *Transactions of the American Geophysics Union* 38: 222–232.

Heat Capacity *See* **Thermal Properties and Processes**

Heat Flow *See* **Thermal Properties and Processes**

HEAVY METALS

D C Adriano, University of Georgia, Aiken, SC, USA
N S Bolan, Massey University, Palmerston North, New Zealand
J Vangronsveld, Limburgs Universitair Centrum, Diepenbeek, Belgium
W W Wenzel, Universität für Bodenkultur, Wien, Austria

Introduction

Certain metals are essential for the normal growth of plants, animals, and humans. In plant nutrition, these are called micronutrients, represented by iron (Fe), copper (Cu), zinc (Zn), manganese (Mn), and molybdenum (Mo). In addition to these micronutrients, arsenic (As), cobalt (Co), chromium (Cr), nickel (Ni), selenium (Se), tin (Sn), and vanadium (V) are essential for animal nutrition. Similarly in human nutrition, the above nutrients are essential, except As and V. They are collectively called micronutrients or trace elements because they are required in only trace amounts (unlike the major nutrients nitrogen, phosphorus, and potassium, required in relatively large amounts) and any excess may result in unfavorable biological response.

Since the advent of the Industrial Age, the environment has been subjected to emission and deposition of anthropogenic chemicals, both organic and inorganic. Consequently, some of these chemicals have accumulated in various ecosystems, both aquatic and terrestrial, to the extent that in numerous instances they have bioaccumulated to exorbitant levels that have the potential to imperil the well-being of biota, including the consumers (i.e., animals and humans). Heavy metals are elements having densities greater than $5\,\mathrm{g\,cm^{-3}}$. The term refers to metals and metalloids that are associated with pollution and toxicity, but also elements that are required by biota at low concentrations. While some of the most prevalent xenobiotic organics have only been released to the environment during the last 3–4 decades (e.g., polyaromatic hydrocarbons (PAHs), dioxin, trichloroethylene (TCE)), several heavy metals have been produced and subsequently emitted for hundreds of years. For example, lead (Pb) was initially produced even before the rise of the Roman Republic and Empire (approx. 2000 years ago).

The ever-increasing production and demand by society for metals indicate the mounting probability of their dispersal and contact with the environment. A metal may be dispersed from the time its ore is mined to the time it becomes a finished product. In some cases the ultimate disposal of the finished product may also lead to metal dispersion. The growing world population and the importance of metals in growing economies just ensures more production of metals and the likelihood of more dispersal, especially in developing countries.

Sources of Metals in the Environment

There are two basic sources of metals input in the environment: natural and anthropogenic. Parent rocks and metallic minerals (or ores) represent the main natural sources. Anthropogenic sources range from agriculture (fertilizers, pesticides, livestock manure, etc.), metallurgy (mining, smelting, metal-finishing, etc.), energy production (leaded gasoline, battery manufacture, etc.), and microelectronics, to waste and scrap disposal. Intensive agricultural practices in industrialized countries, especially in North America and Europe, have been increasing the metal burden in soils with heavy and frequent applications of fertilizers, agrochemicals, and soil amendments. For example, the use of certain rock phosphate in manufacturing phosphate fertilizer could contaminate the soil with cadmium (Cd) – already a concern in New Zealand, Australia, and other countries. Swine manure is typically high in Cu content, i.e., it is used as a growth promotant so using this manure as a soil amendment could increase levels of Cu.

In developed countries such as those in Western Europe and North America, atmospheric deposition plays an important role in contaminating arable soils. In England and Wales, atmospheric deposition represents from 33% (for Cu) to 93% of the total input (for mercury, Hg) of heavy metal contamination. Arable soils received approximately 40% Ni and Zn, and more than 50% Cd, As, and Pb from atmospheric deposition. Cu and Zn are contributed mostly by application of animal manure, Cr by application of industrial by-product waste; Cd and, to a much lesser extent, Cr and Ni are added to the soil by fertilizer and lime application. Sewage sludge represents a minor source of metals in arable soils but, with the banning of ocean disposal and incineration of biosolids, it will assume a substantial role in adding metals to the soil, since disposal of biosolids will be shifted on to the land. In forest ecosystems, soil contamination by metals is virtually all from atmospheric deposition. This distribution of metals in the soils of England and Wales is thought to be typical for most developed countries.

Atmospheric deposition is a transboundary issue that ensures contamination of even pristine areas. Certain metals (e.g., selenium (Se), Hg, As, and antimony (Sb)) are fairly volatile and can be transported over long distances in gaseous form or enriched fine particulates; whereas other metals (e.g., Pb, Cu, Zn, and Fe) are transported only in the enriched particulates. Promulgation of ‘clean air’ policy in many countries may mitigate the impact of atmospheric deposition on the world’s ecosystem, but may also switch the burden to the land, owing to disposal and/or landfilling of waste residues collected from air pollution-control devices or accumulation of waste materials that otherwise could have been incinerated or disposed of in the ocean.

While terrestrial systems serve as the main recipient of metal emission, groundwater and freshwater systems can also be impacted. Metals in the anionic form may leach to the groundwater or be transported laterally to nearby freshwater bodies. Metals known to be mobile because they are negatively charged, i.e., anionic, include the arsenites (As(III)), arsenates (As(V)), selenites (Se(IV)), selenates (Se(VI)), and chromates (Cr(VI)). Freshwater systems can be contaminated via runoff, drainage, or direct emission and/or disposal. The main reservoir for metals in aquatic systems is the sediment. Because of the high buffering capacity (i.e., high alkalinity, salinity, etc.) of marine waters, metals are usually not a concern in this system.

Environmental Contamination

Land contamination with metals has been associated with the accumulation of unwanted waste by-products, either aerially or by application and/or disposal, from various human activities. Chemical degradation of soils is a consequence of the post-Industrial Revolution. The following biological and chemical processes may promote the degradation of multifunctionality of soils and the eventual deterioration of soil quality: weathering, pollution, acidification, organic matter buildup or depletion, nutrient buildup or depletion, and salinization. Other chemical processes in soils, such as ion exchange, complexation, oxidation – reduction, sorption – desorption, and precipitation – dissolution may also influence the extent of its degradation. A chemically degraded soil may offer only limited land use rather than one having a multifunctional role. For example, a site contaminated by Pb from battery storage may not be fit as a playground for children unless remediated.

There are practices that have contributed to chemical degradation of soils: agrochemical application, waste disposal, landfilling and storage of waste, combustion of fossil fuels, mining and smelting of metal ores, application of biosolids and compost, use and storage of petroleum products, deforestation, and dredging. Due to urbanization, industrialization, and intensive agriculture, many types of contamination (‘pollution’ may be a more appropriate term in cases of severe contamination that may result in negative biological response) are recognized as inflicted by gaseous, liquid, and solid forms of metals. Contamination varies from country to country, but generally is more severe in developed countries with high population density. Also, in a given country, the urban areas and their suburbs tend to have more severe contamination.

The importance of metals as probably the most important group of contaminants has been well established. For example, of the 11 most common and dominant contaminants at the US National Priority List sites (the so-called Superfund sites), six (Pb, As, Cr, Cd, Ni, and Zn) are metals. This reflects the extensive production and use of these metals in the USA; similarly, these metals are also the most dominant at hazardous waste sites in affluent countries in Europe and elsewhere. Fortunately, none of these sites include a highly productive arable land; rather, these sites were impacted primarily by the mining and industrial sectors.

Because of enactment of environmental policy limiting the impact and total burden of metals in soil, the US Environmental Protection Agency (EPA) has issued (in 1993) metal limits relative to soil application of municipal sewage sludge on arable land. In addition to maximum cumulative loading, there are also restrictions in ceiling concentrations in sludge and maximum annual loading. For example, in the case of Cd, the permissible maximum annual loading is $1.9\,\mathrm{kg\,ha^{-1}\,year^{-1}}$. The maximum annual loading given for Cr is $150\,\mathrm{kg\,ha^{-1}\,year^{-1}}$. The discrepancy between these two metals (also among the other metals) is due to their different biogeochemical behavior and eventual plant uptake, their difference in phytotoxicity threshold level, and their differing sensitivities to edaphic and environmental factors. The European Commission has also adopted a similar regulatory policy on metal input in arable soils, although generally in a more restrictive manner in terms of loading amounts. There are also differences in limits among countries in the European Union. These transcontinental and interregional variations in limits can be attributed to differences in scientific as well as sociopolitical, economic, and international trade orientation.

Retention and Dynamics of Metals in Soils

In terrestrial ecosystems, the soil is the main repository of metals. Likewise, in aquatic systems, the sediment serves as their ultimate sink. As such, it is very important to understand the various dynamic biogeochemical processes in the soil controlling the transformation, fluxes, and fate of these chemicals. Most importantly, it is paramount to evaluate the extent of the bioavailability (or mobility), more or less on a quantitative basis, to microbiota, plants, animals, and humans as influenced by these various processes. Relevant to this phenomenon is the role of the source term (i.e., form) and chemical speciation of the chemicals of concern. And central to this phenomenon is the role of natural attenuation (NA) in chemical dynamics in soil systems. ('Natural attenuation' refers to the state of equilibrium approaching that of the background, i.e., having negligible risk, induced by indigenous physical, biological, and chemical processes in the soil.)

NA is well established as a remedial strategy for a few organic chemicals, primarily BTEX (benzene, toluene, ethylene, xylene). However, these processes cannot destroy metals, although in some cases they can immobilize them. Natural attenuation can be managed for both organics and inorganics. The kinetics of NA can vary widely between the organics and metals. The important parameters to gauge NA also vary between these two groups of contaminants. For organics (e.g., BTEX), the commonly used parameter is the total concentration (or mass) in the affected plume, and NA, even if accelerated, may take months to years, depending on site and contaminant properties. Also certain organics may degrade into daughter by-products that are more toxic and recalcitrant (e.g., TCE degrades to vinyl chloride); whereas with metals they are generally persistent and immutable, so the 'bioavailable' fraction in the soil should be the most relevant parameter for NA. Unlike common organics, the kinetics of NA for metals, as measured by a decrease in bioavailability, can be relatively much shorter, i.e., in a matter of days.

Metal ions can be retained in soil by the following basic biogeochemical processes: (ad)sorption, precipitation, and complexation reactions. Sorption is defined as the accumulation of matter at the interface between the solid adsorbent and the aqueous phase. This can include ion exchange, formation of surface solid complexes, precipitation, and diffusion into the solid. The lower the metal solution concentration and the more sites available for adsorption, the more likely that adsorption process will determine the soil solution concentration. Once deposited in the soil, metals interact with the soil minerals and organic constituents. However, the fate of metals in the soil environment is dependent on both soil properties and environmental factors.

Adsorption

Charged solute species (ions) are attracted to the charged soil surface by electrostatic attraction and/or through the formation of specific bonds. Retention of charged solutes by charged surface is broadly grouped into specific and nonspecific retention. In general terms, nonspecific adsorption is a process in which the charge on the ions balances the charge on the soil particles through electrostatic attraction; whereas specific adsorption involves chemical bond formation between the ions and the sorption sites on the soil surface. If a nonspecific adsorption process solely controls metal adsorption, then the adsorption capacity of the soil is dictated by its cation exchange capacity (CEC). However, in many soils the amount of metal sorbed exceeds the CEC of the soil. This infers that, in addition to nonspecific adsorption, other processes such as specific adsorption, precipitation, and complex formation also contribute to the overall retention.

Both soil properties and soil solution composition determine the equilibrium of metals between solution and solid phases. The concentration of metals in soil solution is largely influenced by the pH and the nature of both organic and inorganic anions. The effect of pH values greater than 6 in lowering free metal ion activities in soils can be attributed to the increase in pH-dependent surface charge on oxides of Fe, Al, and Mn, chelation by organic matter, or precipitation of metal hydroxides.

There are three reasons for the effect of inorganic anions on the adsorption of metals: (1) inorganic anions form ion-pair complexes with metals, thereby reducing their adsorption on to soil particles; (2) the specific adsorption of ligand anions is likely to increase the negative charge on soil particles, thereby increasing the adsorption of metal cations; and (3) specifically sorbed anions (e.g., phosphate) strongly compete with metal anions such as arsenate and selenates, resulting in their desorption.

Complexation

Metals form both inorganic and organic complexes with a range of solutes in soils. For illustration, the effect of inorganic anionic complex formation on the adsorption of Cd^{2+} by soils is discussed. Generally, chloride can be expected to form a soluble complex with Cd^{2+} as $CdCl^{+}$, thereby decreasing the adsorption of Cd^{2+} on to soil particles. In contrast to inorganic ligand ions, Cd^{2+} adsorption by kaolinite, a variable-charge mineral, could be enhanced by the presence of organic matter via the formation of an adsorbed organic layer on the clay surface. As might be expected, the organic component of soil constituents has a high affinity for metal cations because of the presence of ligands or groups that can chelate metals. With increasing pH, the carboxyl, phenolic, alcoholic, and carbonyl functional groups in soil organic matter dissociate, thereby increasing the affinity of ligand ions for metal cations. The general order of affinity for metal cations complexed by organic matter is as follows:

$$Cu^{2+} > Cd^{2+} > Fe^{2+} > Pb^{2+} > Ni^{2+} > Co^{2+} > Mn^{2+} > Zn^{2+}$$

Precipitation

Precipitation appears to be the predominant process of metal immobilization in alkaline soils in the presence of anions such as carbonate, hydroxide, and phosphate, especially when the concentration of metal ion is high. Coprecipitation of metals, especially in the presence of iron oxyhydroxide, can also be expected, and often such interactions lead to significant changes in the surface chemical properties of the substrate. Precipitation as metal phosphates is considered to be one of the mechanisms for the phosphate-induced immobilization of heavy metals, especially in substrates containing high concentrations of metals.

Liming typically enhances the retention of metals; for example, increased retention of Cr(III), Cd, and Pb with a lime-induced increase in pH can be anticipated. The pH of the lime-treated soil, which ranges from 7 to 8, coincides with the effective precipitation range for a number of metals as their respective metal hydroxides (for example, Cr(III) as $Cr(OH)_3$). An increase in pH due to liming also increases the negative charge of variable-charge soils, enhancing metal adsorption.

Bioavailability–Natural Attenuation Interactions

Since most metal retention in soils occurs in the upper layer of the soil profile (especially those deposited aerially), usually involving a large tract of land, a possible landscape scenario depicts a vegetated system, unless the metal levels in soils are phytotoxic. Otherwise, for soil conservation and environmental protection purposes, highly metal-contaminated soils need to be amended as an integral part of a remediation scheme. Thus plants play a central role in the bioavailability of the metals in question and their NA. We hypothesize that plants, directly or indirectly, heavily influence the degree and kinetics of NA and modify the influence of certain soil amendments when employed to accelerate NA.

A more generic definition of 'bioavailability' is the potential for living organisms to take up chemicals from food (i.e., orally) or from the ambient environment (i.e., externally) to the extent that the chemicals may become involved in the metabolism of the organism. More specifically, 'bioavailability' refers to the biologically available chemical fraction (or pool) that can be taken up by an organism and can react with its metabolic machinery; or it refers to the fraction of the total chemical that can interact with a biological target, for example a plant. In order to be bioavailable, metals have to come in contact with the plant (i.e., chemical accessibility) to be able to enter a plant root. In essence, for a metal to be bioavailable, it has to be mobile, transported to the plant, and be in an accessible form. Hence, the bioavailability (more specifically, 'phytoavailability') of a metal in soil may be defined by the amount of metal absorbed by growing plants or by concentration in the harvested plant tissue. Plants typically absorb much less than 1% of the metals present in the soil. This expression does not indicate the extent to which metals present in soils are bioavailable nor the duration of bioavailability. Therefore, bioavailability of metals must account for not only the plant uptake in one growing season, but also the total amounts potentially available over time. The concepts of NA and bioavailability are being employed increasingly in remediation strategies.

The bioavailable fraction can be particularly important for metals, despite its dynamic nature and sensitivity to edaphic and environmental factors. (The 'bioavailable' fraction of metals in soil can be estimated by an extraction procedure using weak electrolytes; such a procedure has been used successfully by soil chemists for decades for micronutrient metal (Fe, Cu, Zn, and Mn) requirement by crops.) While the total content of metals is still being widely used for regulatory purposes, there has been a significant move, especially in Western Europe, to adopt the bioavailable fraction in risk assessment.

It is important to emphasize the critical role of source term (or form) and chemical speciation of metals in bioavailability. The 'source term' refers to the physicochemical form of the metal after it is released and deposited in soil. For example, Pb bound in metal ores in mining areas has very low bioavailability to humans compared with Pb bound to dust in urban areas. Thus, greater exposure of children to Pb in big cities can be expected. Speciation refers to the ionic or complexed (i.e., with an ion pair) condition of a particular metal within a given oxidation (i.e., valency) state. There is overwhelming evidence that certain chemical species (rather than total concentration) of metals in the soil solution are more bioavailable to plant roots than others. In microbiology, the free-ion activity model has been proposed as a general hypothesis, claiming that metal uptake by microbes is largely related to the activity of free ions in solution. While this model may apply to microbial uptake, it may not be so with higher plants. Indeed, metal–chelate (ethylenediaminetetraacetic acid, EDTA, or diethylenetriaminepentaacetate, DTPA) complexes have been shown to facilitate phytoextraction of metals, in this form, by *Brassica juncea* (Indian mustard). Furthermore, certain inorganic ligands may facilitate uptake. For example, Cd–chloro complexes are potentially more bioaccumulative, as demonstrated with the field potato in Australia. Thus it is evident that free ions or complexed ions should be favored over the total metal concentration in soil solution as a predictive tool for bioavailability (or, conversely, NA).

Chemical extraction of soil is another tool widely used to assess bioavailability of metals. Selective extractants using dilute acids (e.g., HCl, NH_4OAc) have been extensively used in soil-testing as a diagnostic tool for micronutrient deficiency. On the other hand, sequential extraction provides more information, in addition to the bioavailable pool, on the redistribution of the metal amongst organic and mineral soil constituents. When amended, it can be expected that the majority of the metals in the soil would be sequestered in less labile pools represented by the organic matter, Fe and Mn oxides, and silicate clays.

Quite often plant performance (yield, chlorosis and/or necrosis, root morphology), plant quality (total concentration of metals, nutrients, etc.), plant diversity, and metal-induced biochemical stress responses in plant tissue are used as direct indicators of remediation efficacy. Green, lush vegetation in a highly contaminated soil is a good sign of healthy soil, indicating favorable soil–microbe–plant interrelations. Other assessment tools include bacterial biosensors that detect the bioavailable fraction of a certain metal or suite of metals. For effective monitoring it is often necessary to deploy a mix of these tools.

Biological Response to Metals

When metals exceed certain threshold levels in soils, the organism being impacted may exhibit a negative biological response. The target organisms involved range from microorganisms, small invertebrate animals and other microfauna, and microflora to higher plants, including crops. Either one of these organisms or a combination of them may be used to assess their behavior, which has become a pivotal part of ecologic risk assessment. When their behavior becomes detrimentally affected, impacting on their normal physiological and/or biochemical functions, the metal can be suspected to pose some risk. These alterations in their usual biological functions could lead to disruption in community and/or population structure (i.e., less biodiversity), eventually leading to death. Movement of high concentration of metals in solute affecting groundwater quality is also a form, although not biological, of risk.

Likewise, humans could be at some risk when unnecessarily exposed to metals. The risk depends on the exposure route and the dose. The total dose represents the sum of specific doses resulting from each exposure pathway:

$$\begin{aligned}\mathrm{Dose}_{\mathrm{total}} = {} & \mathrm{Dose}_{\mathrm{soil}} + \mathrm{Dose}_{\mathrm{food}} + \mathrm{Dose}_{\mathrm{water}} \\ & + \mathrm{Dose}_{\mathrm{air}} + \mathrm{Dose}_{\mathrm{dermal}}\end{aligned}$$

with units in milligrams of metal per kilogram of body weight per day.

The first three categories of dose represent ingestion of soil and/or dust (especially in children), edible food, and water, the fourth one representing inhalation of polluted air and the last representing skin absorption from polluted soil, water, or air. The dose relationship above indicates that young children, because of their low body mass, are the most sensitive segment of the population. Industrial hygiene is necessary to limit exposure of workers to metals via inhalation (e.g., vapor and/or fine dust) and/or skin contact (e.g., Ni). The relationship above also indicates that the most common exposure pathway for the general population is via ingestion of food and drinking water. The main concern in the former is the consumption of vegetables high in Cd and seafood high in Hg. Drinking well water high in As is of concern in some parts of the world, particularly in Bangladesh and West Bengal, India, where millions of people are at risk.

Just like in ecological risk, metals can affect humans in a variety of ways – depending on the type, form, and chemical speciation of the metals, age and body weight, dietary and/or nutritional status, exposure route and dose, etc. Because metals typically exist as metallothioneins, i.e., metals binding strongly to proteins, when in excess, they can disrupt the functioning of metalloenzymes. Such dysfunctions can eventually be manifested by disorders, for example in the central nervous system (by Hg and Pb), kidney (by Cd), liver (by As), blood (by Pb), and skin and extremities (by As). Under acute exposure, multiorgans can be affected, which can eventually lead to death.

Soil Remediation

The use of conventional, engineering-type techniques to clean up metal-contaminated soils is typically invasive and expensive. For example, in soil-washing, the soil has to be excavated first and usually treated off-site. The treated soil may be returned to the site or landfilled. The washing not only irreversibly destroys the integrity of the soil in a general sense, but also produces a secondary waste – the washing effluent, which is usually a chelate solution now laden with metals.

The most logical, ecologically friendly and economically viable method is by *in situ* immobilization of metals using abundant, inexpensive soil amendments. This technique is particularly suitable for the treatment of big tracts of land where the soil is mostly surficially contaminated. Soil amendments that have been successfully tested in North America, Europe, and Australia include liming materials, rock phosphate (hydroxyapatite), alkaline biosolids and/or compost, Fe-rich by-products (steel shot, red mud, 'Fe-rich' from aluminum processing), and coal residues (alkaline coal fly ash, beringite).

Traditionally, the use of unenhanced (or noninvasive) natural processes as part of a site-remediation strategy is called natural attenuation. Thus the US-EPA coined the term 'monitored natural attenuation' (MNA) when NA is employed within the context of a carefully controlled and monitored site cleanup strategy to be able to achieve site-specific remediation objectives within a time frame that is more reasonable than that offered by other, more invasive methods. A new twist to the definition has been proposed, that natural processes can be accelerated by adding to the soil inexpensive amendments that are cost-effective and minimally invasive. These materials can be incorporated into contaminated soils to speed up and optimize, in the case of metals, their immobilization as mediated by key processes such as (ad)sorption, precipitation, complexation, and redox reactions. The extent of immobilization can be measured by K_d (partition coefficient), which also serves as an index for bioavailability. A high K_d means very strong binding to the soil solid phase, rendering very low concentrations of the metal in the solution phase. Thus, a high K_d translates to low bioavailability induced by effective immobilization.

The 'driver' processes indicate that, when they are dominantly operative in the soil system, the K_d is high and bioavailability is low (or mobility and leaching potential are also low). In order to sustain the immobility (or high K_d) of metals, the 'counter' processes (e.g., desorption, dissolution, high oxidation potential) should be minimized or prevented, otherwise they can impede metal immobilization processes. For example, precipitation as metal phosphate has been shown to be a major mechanism for the immobilization of metals, such as Pb and Zn by phosphate compounds. These new compounds have extremely low solubility over a wide pH range, rendering phosphate application an attractive technology for managing metal-contaminated soils. The formation of the new solid phase (i.e., precipitates) occurs when the ionic product in the solution exceeds the solubility product of that phase. Recent studies using X-ray absorption fine structure (XAFS) indicate that formation of surface precipitates may occur even when solution concentration is undersaturated with respect to homogeneous precipitation of the pure metal-precipitate phase. In normal soils, precipitation of metals is unlikely, but in highly metal-contaminated soils, this process can play a major role in immobilizing metals.

The ability of apatite to immobilize dissolved Pb^{2+} as precipitates is more commonly manifested as hydroxypyromorphite or as chloropyromorphite. Two processes for the reaction of dissolved Pb^{2+} with apatite are proposed: first, Pb can react with apatite through hydroxyapatite (HA, $Ca_{10}(PO_4)_6(OH)_2$); second, Pb^{2+} can substitute for Ca^{2+} in the apatite (eqn 3). Thus, (Ca, Pb) apatite could be potentially formed by adsorption of Pb or by dissolution of HA, followed by coprecipitation of mixed apatites. Dissolution of apatite (eqn 1) is an important initial step in the immobilization of Pb as pyromorphite.

$$Ca_{10}(PO_4)_6(OH)_2(s) + 14H^+(aq) \rightarrow 10Ca^{2+}(aq) + 6H_2PO_4^-(aq) + 2H_2O \quad [1]$$

$$10Pb^{2+}(aq) + 6H_2PO_4^-(aq) + 2H_2O \rightarrow Pb_{10}(PO_4)_6(OH)_2(s) \quad [2]$$

$$Ca_{10}(PO_4)_6(OH)_2(s) + xPb^{2+} \rightarrow (Ca_{(10-x)}Pb_x)(PO_4)_6(OH)_2(s) + xCa^{2+} \quad [3]$$

The immobilization of As, Cd, Pb, and Zn in smelter-contaminated soils using diammonium phosphate (DAP) is also possible. Application of high levels of DAP at a rate of 2300 mg P kg^{-1} could be very effective for immobilizing Cd, Pb, and Zn in the contaminated soil. Activity–ratio diagrams indicate that DAP decreases solution concentration of these metals by forming metal-phosphate precipitates with low solubility products.

Similarly, since bioavailability of metals to plants is typically greater in acidic than alkaline soils, neutralizing agents in the form of lime are usually added to acidic soils. Although the primary incentive in liming acidic arable soils is the suppression of toxic bioavailable Al and Mn to plants, liming is increasingly being practiced as a management tool to immobilize metals in soils, biosolids, and mine tailings, thereby reducing their bioavailability for plant uptake and transport to groundwater. In fact, alkalinization of biosolids is commonly done to enhance their immobilization potential for metals. Several reasons have been attributed to the lime-induced immobilization of metals: increases in negative charge (CEC) in variable-charge soils; formation of strongly bound hydroxyl metal species; precipitation of metals as hydroxides; and sequestration due to enhanced microbial activity. A range of liming materials are available, which vary in their ability to neutralize the acidity and immobilize metals: calcite ($CaCO_3$), burnt lime (CaO), slaked lime ($Ca(OH)_2$), dolomite ($CaMg(CO_3)_2$), and slag ($CaSiO_3$).

Since the main risk in highly contaminated soils arises from the mobility (transport) and bioavailability (uptake), and subsequent accumulation (biotoxicity) in biota, any attempt to remediate such sites should effectively control these two most important parameters. Accordingly, these two processes should be monitored by a battery of tests including, but not limited to, chemical speciation, selective and sequential extraction, microbial tests, plant performance, and quality.

Acknowledgment

Award number DE-FC09-96SR18546 from the US Department of Energy to the University of Georgia Research Foundation is gratefully acknowledged.

Further Reading

Adriano DC (2001) *Trace Elements in Terrestrial Environments: Biogeochemistry, Bioavailability and Risks of Metals*, 2nd edn. New York: Springer.

Bolan NS, Adriano DC, and Naidu R (2003) Role of phosphorus on (im)mobilization and bioavailability of heavy metals in soil-plant system. *Review of Environmental Contamination and Toxicology* 177: 1–44.

Bolan NS, Adriano DC, and Curtin D (2003) Soil acidification and liming interactions with nutrient and heavy metal transformation and bioavailability. *Advances in Agronomy* 72: 215–272.

Deutsches Institut für Normung (1995) *Soil Quality Extraction of Trace Elements with Ammonium Nitrate Solution*. DIN 19730. Berlin, Germany: Beuth-Verlag.

Lombi E, Wenzel WW, and Sletten RS (1999) Arsenic adsorption by soils and iron-oxide-coated sand: kinetics and reversibility. *Journal of Plant Nutrition and Soil Science* 162: 451–456.

Ross SM (ed.) (1994) *Toxic Metals in Soil–Plant Systems*. New York: John Wiley.

Sarret G, Vangronsveld J, Manceau A *et al.* (2001) Accumulation forms of Zn and Pb in *Phaseolus vulgaris* in the presence and absence of EDTA. *Environmental Science and Technology* 35: 2854–2859.

Vangronsveld J and Cunningham SD (1998) *Metal-Contaminated Soils: In Situ Inactivation and Phytorestoration*, pp. 265. Berlin, Germany: Springer-Verlag.

Vangronsveld J, Ruttens A, Mench M *et al.* (2002) *In situ* inactivation and phytoremediation of metal/metalloid contaminated soils: field experiments. In: Wise DL, Trantolo DJ, Cichon EJ, Inyang HI, and Stottmeister U (eds) *Bioremediation of Contaminated Soils*, 2nd edn, pp. 859–884. New York: Marcel Dekker.

Wenzel WW, Lombi E, and Adriano DC (1999) Biogeochemical processes in the rhizosphere: role in phytoremediation of metal-polluted soils. In: Prasad NMV and Hagemeyer J (eds) *Heavy Metal Stress in Plants – From Molecules to Ecosystems*, pp. 273–303. Heidelberg, Germany: Springer-Verlag.

HILGARD, EUGENE WOLDEMAR

R Amundson, University of California–Berkeley, Berkeley, CA, USA

The field of soil science emerged during the great scientific advances of the nineteenth century. During this formative period, it is likely that no one in the USA cast as large an intellectual and personal shadow as Eugene Hilgard. Hilgard's initial intellectual contribution to pedology occurred in his report on the geology and agriculture of the state of Mississippi, a report which presented an erudite and novel synopsis of soil formation and the soil geography of Mississippi. Hilgard's interests gradually shifted from his initial geological efforts toward soils and agriculture, and this growing interest in soils followed his academic transfer to Michigan and then ultimately to Berkeley, where the second half of his long career was dedicated to a great expansion of soil science education and research at both a state and a national level. His accomplishments and experiences seem to surpass the 83 years in which he lived, and they continue to serve as a standard with which modern achievements and accomplishments in the natural and agricultural sciences must be compared.

Hilgard was born in Zweibrücken, Germany, on 5 January 1833. His father, Theodore Erasmus Hilgard, was a prominent lawyer and served as Chief Justice of the Court of Appeals in his province. Hilgard's mother, Margaretha Pauli, was the descendant of generations of clergy, many of great importance to royalty and universities in France and Germany. Hilgard's parents possessed liberal views, and his father resigned his court appointment rather than adopt a revised and less progressive legal code. This resignation and the changing political landscape in Germany led to the Hilgard family's decision to emigrate to America. In 1836, the Hilgard family arrived in Belleville, Illinois, a settlement already home to an enclave of cultured and educated German families that had emigrated to the USA for political reasons.

Life in the Hilgard household of Belleville probably bore little resemblance to that of average pioneer families of the time. Eugene was the youngest of nine children, and with local schools being considered unsatisfactory by the family, Hilgard's father undertook the education of the children himself. Emphasis on mathematics and language was particularly strong, and, being educated in largely French institutions, Hilgard's father imposed two 'French only' days at the home each week (forcing the young boys to retreat outdoors to speak German). The children were encouraged to study the local botany and natural history, and this effort was greatly aided by the family's rich library collection. The breadth and rigor of this early education are made even more impressive by the fact that, in subsequent studies at universities or technical institutes, Hilgard was commonly far ahead of his peer group in academic preparation and abilities.

Although Belleville offered many opportunities and experiences, mosquitoes bore typhoid and malaria, the latter of which is believed to have caused fevers and sickness which plagued him for much of the remainder of his life. Malaria is also believed to have contributed to a weakening of his eyes, which also posed a lifelong handicap. At the age of 15, suffering from malarial maladies and greatly missing his older brother Theodore, who had left to study medicine in Europe, Hilgard's father let him travel with his brother Julius to Washington, DC, where Julius was an assistant in the US Coast Survey. Hilgard's time in Washington, and a later visit to Philadelphia, provided the young Hilgard with a variety of intellectual and scientific opportunities. Hilgard's growing interest in chemistry, and the few opportunities to study this emerging field in America, led Hilgard to travel to Heidelberg to study, where his brother Theodore was already a medical student.

At Heidelberg, Hilgard was an eager student, engaging in laboratory experiments and attending lectures. However, many of the courses were found to be, in Hilgard's view, uninspiring, and the social code was found to be unacceptable and disappointing. These deficiencies, combined with political unrest in Germany, led Hilgard and his brother to travel to Zurich. In Zurich, Hilgard spent three semesters studying botany and geology, and, due to the rigor of his preparatory education, was soon selected to be an assistant in chemistry both in lectures and in the laboratory. Following this brief stay in Zurich, Hilgard moved to the Mining Academy at Freiburg for advanced study in mining and metallurgy. The bleak physical surroundings, unsocial conditions, and physical hazards at Freiburg ultimately forced Hilgard's departure. The final blow came when Hilgard was overcome by mercury vapor during a distillation process, causing Hilgard's physician brother to declare that Eugene should follow the outdoor life of a geologist or botanist, medical advice he provided consistently on multiple occasions during later years.

Leaving Freiburg, and after engaging in recuperative exercise in the Swiss countryside, Hilgard returned to Heidelberg, this time to more advantageous academic prospects. He was soon drawn to the recently hired Professor Robert Bunsen (of Bunsen burner fame), whose research interests spanned geology to chemistry, much as Hilgard's interests did. Bunsen encouraged Hilgard to choose, as his doctoral thesis, an enquiry into the chemistry of the four components of a burner flame. Hilgard produced a definitive dissertation on the chemistry of the gases in the flame components and, on 3 October 1853, received a degree that was bestowed with the university's highest honor, *summa cum laude*. Hilgard was then only 20 years old and, amazingly, had not received any formal degrees of any type prior to that point.

Following his dissertation examination, Hilgard suffered a relapse in his health, and was advised to travel to a suitable climate, this time Spain. While there, he engaged himself by learning the language and customs, and his love of music led to his introduction to the talented daughter of a retired Colonel, Jesusa Bello, who in 1860 became his wife. In the spring of 1855, his health improved, Hilgard obtained passage to New York and then to Washington, where, for lack of other suitable employment, he continued his chemical research at the Smithsonian Institution. Soon, Hilgard received an offer to be the assistant to the State Geologist of Mississippi, a position in which he was adequately forewarned not only of the difficult director then in charge, but also of the lack of scientific challenges in the Paleozoic formations of the state. But Hilgard accepted the position, setting into motion challenges which served to test his intellect, creativity, fortitude, and political talents – an arsenal of skills which served him well for nearly 50 more years.

In retrospect, one cannot envision a more unlikely launching site for a such an impressive career, or for many of the initial conceptual and quantitative developments in soil science. As Fredrick Slate wrote, in the National Academy's biographic memoir of Hilgard, "In all these relations, it seems probable that Hilgard's training, because it was foreign, must have helped him to a certain detachment that would favor freshness in attack and in conception of aims." Indeed, Hilgard's supervisor in Mississippi proved to be ineffective, if not incompetent, and soon left the position in 1857. Hilgard used this intervening period to his advantage, meeting and consulting with the geologists of the neighboring states of Kentucky, Arkansas, and Alabama. Notably, Dr. David Dale Owen, of Indiana, provided Hilgard with the advice to place emphasis on the soils and agriculture of the state, advice that Hilgard clearly seized. Following a trip to Washington in 1857, to escape the intolerable conditions under the former Head Geologist, Hilgard returned to Mississippi and was himself appointed State Geologist.

Only 25 years old, and for the most part intellectually isolated, Hilgard set into motion his plan for the geological investigation of the state. Legislative support and approval of his work were initially almost withdrawn, but political skirmishing succeeded in reestablishing the survey and providing Hilgard the time to complete his report. The completed document was published in 1860, but was not released until 1866 due to the onset of the Civil War.

The 1860 report is remarkable in that it, and subsequent investigations, established Hilgard as a preeminent scholar in the already established field of geology, later earning him the praise of being the 'father of the American Tertiary.' Hilgard's files of correspondence contain letters from, among others, Louis Agassiz and the father of geology himself, Sir Charles Lyell. Hilgard's work in Mississippi earned him the later opportunities to explore and clarify the geology of Louisiana, and the 'Mississippi Embayment,' and many of his early publications center on a variety of geological and paleontological problems of the then 'Southwest.' This eminent standing in the geological community was later important in Hilgard's efforts to establish a federal program of soil survey.

However, even more innovative was the second half of the 1860 report, which focused on the principles of agriculture and the agricultural features of the state. The first portion, 'Principles of Rational Agriculture,' is truly remarkable. It begins, most interestingly, with a question: 'What is a soil?' Hilgard answers it, and addresses soil's formation and relationship to plant production, in a most rigorous and enlightening discussion. It is important that Hilgard posed the question using the article 'a,' which explicitly indicates that he recognized that soils are naturally varying entities that, in a given location, possess characteristic features. In Hilgard's mind, at the time, soil (in the present sense) contained 'soil,' or what we might now call the A horizon, and 'subsoil,' the material lying below the 'soil.' Though not expressing it in these terms, Hilgard therefore recognized the 'profile' and 'horizon' concept. More importantly, he presented a revealing discussion of soil biogeochemistry, including a discourse on the oxidation and reduction of iron compounds, and included a discussion on the formation and role of humus in agricultural practices.

Published in 1860, the report sat in storage until the end of the war. During this period, Hilgard retained his position as State Geologist, and (as a sign of the regard in which he was held by the Legislature) he continued to receive a salary of $1250 per year and an account to buy chemicals and carry out analyses. He also served the Confederacy by exploring for nitrates, and he attempted to install calcium lights on the bluffs above Vicksburg to illuminate federal gunboats – a plan that ultimately failed. He was also placed in charge of university property during the war, maintaining the integrity of the campus through the hostilities and during periods when it served as a hospital for the wounded.

In 1866, due to political insecurities, Hilgard joined the university as a professor of chemistry, the beginning of his long and varied academic career, one which marked a continuous shift from a primarily geological focus (though he maintained such an interest to the end) to one in soils and agriculture. In 1871, his professorial title was changed to Professor of Experimental and Agricultural Chemistry. This same year, Hilgard was the delegate from his university to the first convention of agricultural colleges established under the Morrill Act of 1862. While there, Hilgard first met D.C. Gilman, who at that time was at Yale. A few years later, it would be Gilman, then President of the fledgling University of California, who would induce Hilgard to move to Berkeley.

However, in the meantime, Mississippi moved to expand academic programs and instruction in agriculture. A small student enrollment, and political opposition and threats to separate the Agricultural College from the remainder of the campus, ultimately seemed to drain Hilgard and, in 1873, he accepted an offer, that had twice before been extended, to become Professor of Geology and Natural History at the University of Michigan in Ann Arbor.

While Ann Arbor, reportedly the premier public institution at that time, offered the cultured Hilgard family (which by that time included three children) a stimulating atmosphere, a number of issues forced Hilgard to look far beyond Michigan for his future. First, he was obligated to teach large classes, leaving him little time for original research. Second, the agricultural components of the Michigan educational system were located in Lansing, a situation which Hilgard unsuccessfully lobbied against. And finally, to a family of southerners (and a wife from Spain), the Michigan winters proved to be unbearable. In 1874, while attending a meeting of the American Association in Hartford, Connecticut, Hilgard again met Gilman, who was now President of the University of California, and who was also looking for a suitable professor of agriculture (after recently dismissing the first occupant of this position). It is reported that Gilman immediately offered the Chair to Hilgard, but Hilgard demurred. Gilman then succeeded in inducing Hilgard to travel with him back to California and to take a leave of absence from Michigan. This request was, as might be expected, met with strong objections by President Angell of Michigan, who "warned strenuously against going to such a hornet's nest as the university at Berkeley was reputed to be." None the less, Hilgard took leave, met Gilman in Chicago, and crossed the continent in the company of the president of the new University of California. Upon his arrival in Berkeley, Hilgard delivered lectures over a 6-week period on 'The Origin, Properties, and Functions of Soils,' which was attended by a relatively large group of students, many of them graduates.

While all is reported to have gone well during the visit, Hilgard returned to Ann Arbor "without a lingering thought of returning to California." However, yet another brutal winter and a medical relapse caused Hilgard to change his mind and, in 1875, he accepted Berkeley's offer. As would be expected, the Michigan faculty did their best to prevent the the loss of Hilgard (who was elected to the National Academy of Sciences in 1873). President Angell even promised to keep Hilgard's chair open for 6 months should he eventually change his mind.

Upon his arrival in Berkeley, it certainly must have entered Hilgard's mind to return directly to Michigan. His recollection of that period, and his response to it, summarizes the monumental struggle, and cultural sacrifices, that Hilgard subjected himself and his family to in what was now the second half of his career:

> When, after an uneventful overland voyage, we landed in Oakland, I found to my intense disappointment that Dr. Gilman was on the point of leaving the University of California for the presidency of Johns Hopkins University, at Baltimore. Had I known this, I might have never left Ann Arbor, and I thought for a moment of turning round immediately. But Dr. Gilman protested that

> I ought at least to try and study the situation; and, although he was leaving, I would have the strong help of the Board of Regents, and a new and fruitful field of investigation instead of the mere routine of teaching in Ann Arbor. And he gave me as his farewell a numerously attended reception at which he introduced me to a large number of influential people with warm recommendations for aid in my task. So I did not return to Ann Arbor, but resolved to fight it out as I had done in Mississippi before. The fight turned out to be quite as hard as anticipated, for a number of years. But I won in the end; and the climatic advantages proved for myself and my family an ample offset in the prolongation of life and health to the social pleasantness we had relinquished in making the change.

The cultural and academic contrasts between Berkeley and Ann Arbor at the time must have been severe. Photographs of the Berkeley campus at the time show a few isolated university structures in the midst of what appears to be pastureland. Streets were consistently reported to be rutted and muddy, and cattle and coyotes were reported to roam across the campus. To a man whose command of language (speaking German, Spanish, English, French; reading Sanskrit, Italian, Greek, Latin, and Portuguese) was second only to the university's Professor of German, this had to have been a dismal turn of events.

The Berkeley phase of Hilgard's career, beginning in 1875 as Professor of Agriculture and Botany, began with numerous challenges to his and the fledgling College of Agriculture's future. Internally, it is reported that suspicion by his colleagues over the role of agriculture in the university posed a challenge, one that was diffused by Hilgard's command of languages and literature. More significantly, opposition by the state's agricultural constituency, many of whom promoted the concept of a more practical, as opposed to scientific, approach to agricultural education, was a serious issue. Hilgard diffused this opposition head on, through a series of public lectures and discussions with the agriculturalists of the state. His efforts on public outreach and education seem extraordinary to the modern academician and certainly give a sense of the provincial nature of the state at that time and to the tenuous foothold that agricultural education held. Hilgard never ceased his outreach and educational efforts, writing and lobbying in newspapers and the general press for the remainder of his career.

However, to the scientific and international community, it was Hilgard's reinvigorated research, stimulated by new circumstances and climates, that is the hallmark of this period. It is challenging, and ultimately a disservice, to try to summarize his accomplishments in such a brief fashion. Hilgard's bibliography lists 229 publications: books (including his famous *Soils* of 1906), multivolume reports, newspaper and magazine articles, an abundance of research articles in the premier journals such as *Science, American Journal of Science*, etc. – on topics ranging from geology, soils, and plant pathology to education and language. In addition, due to his poor eyesight, Hilgard possessed a typewriter as early as 1876. As a result of his scientific notoriety, his records include carbon copies of more than 20 000 letters to people ranging from local farmers to Sir Charles Lyell. The range of his intellectual prowess, and his sheer fortitude to maintain scholarly output and contact, is staggering even to those now accustomed to various means of computer- and electronic-aided correspondence and publishing.

Hilgard's biographers have consistently emphasized several key achievements during his Berkeley era, which will be highlighted here. First, Hilgard accepted the commission by the federal government to conduct and publish the Cotton Census for the southern states and California (1884). The commission provided him with $25 000 to hire staff, collect samples, conduct analyses of cotton and, most importantly, soils – putting Hilgard's developing College of Agriculture on the national stage (and, of additional importance, at no cost to the state of California). On the surface this may appear an unlikely undertaking for someone of Hilgard's background, but a perusal of the California report reveals Hilgard's intent. The report is primarily the first synthesis of the soil geography of the state, complete with regional discussions and chemical analyses. Included also is the first 'Agricultural Map' of the state (outlining the geographical/soil provinces in a way still used today).

The second contribution mentioned here is Hilgard's quantitative attack on the origin and remediation of soil salinity and sodicity. California, like many arid and semiarid regions, contains soils in certain geographical settings that naturally possess high levels of salt. These are also areas which, due to irrigation, are susceptible to human-induced expansions of salinity due to poor irrigation management. Remarkably, Hilgard (using his nineteenth-century chemical technology) was able to decipher the general framework of the related, but differing, roles of various salts and sodium on crop production and water infiltration. He was able to determine that the application of gypsum would prove to be an effective means of remediating sodic soils, and thereby improving structure and hydraulic characteristics. He also, as part of this work, perceptively wrote that irrigation systems without adequate drainage are doomed to waterlogging and salinization, a principle that has yet to be embraced by the state of California's government and its citizenry. The work became widely disseminated around the world, in regions, arid or not, where salt problems hindered agricultural productivity.

Hilgard's displacement to the semiarid California, and his summer expeditions in the arid Pacific Northwest as a consultant for his railroad 'king' cousin Henry Villard (on the Northern Transcontinental Railroad Survey), provided him with insights into arid soils. These trips provided a stark contrast to his soil experiences in Mississippi. Thus, he developed a clearly conceived comparison of soils and soil-forming processes in the arid versus humid regions of the country, focusing in particular on carbonate and plant essential nutrients.

Among Hilgard's great efforts at Berkeley is an equally important 'near miss' in an activity of far-ranging importance and intrigue. Hilgard's experiences in Mississippi and during the Cotton Census effort sharpened his view that the nation needed an organized program of soil (or, using his term, 'agricultural') survey. His reading of the charter of the newly formed US Geological Survey (USGS; one of the first, if not the first, federally funded scientific programs) revealed that one of its missions was the 'classification of public lands,' which Hilgard viewed as including soils. Hilgard's attention on the USGS coincided with the selection of John Wesley Powell as its director (1881). Powell, like Hilgard, possessed broad-ranging interests, and additionally shared Hilgard's concern over the distribution and use of land. Powell's tenure as USGS director was marked by great highs and lows and, while agricultural survey was but a small part of his overall concern, he responded to Hilgard's (sometimes almost annoying) lobbying by offering him a position in the USGS to undertake the agricultural survey program. Hilgard, now fully entrenched at Berkeley and in the agricultural arena, declined this overture, but continued to lobby his eastern colleagues (many with powerful political connections) to push Powell to engage in agricultural pursuits.

In 1888, the Agriculture Department was elevated to a cabinet-level department, and Hilgard was nominated (and approved by Congress) to be the Assistant Secretary in charge of scientific investigations. This would have given Hilgard the opportunity personally to formulate a federal survey program. Additionally, the new US Department of Agriculture (USDA) appealed to Powell as a more stable funding base for the USGS, which then as now was part of the Department of Interior. To the shock and dismay of his eastern colleagues, Hilgard turned down the USDA position. The reasons that Hilgard cited for declining included issues which by then were well known to those familiar with his writings: the unfavorable climate of Washington and the lack of financial compensation (which he struggled to maintain for both himself and his research programs throughout his career). Additionally, one senses, in reviewing his correspondence, that he was hesitant to relinquish the fame and stature he enjoyed in Berkeley and in California (he was well acquainted with senators, governors, and business leaders in the region).

This decision proved a disaster for Hilgard, and produced ramifications for soil survey and the earth sciences in the USA that continue today. First, his friends and colleagues were embarrassed by his decision (since the congress and president had already been convinced on Hilgard). Second, Powell, lacking personal support within the USDA, never succeeded in transferring the USGS to the USDA and he himself was then subjected to a series of continuing political attacks which eventually forced his retirement. Third, while soil survey did indeed develop in the USDA (1899), it was headed by a young and ambitious scientist named Milton Whitney. Early correspondence between the two men reveals that Hilgard was amused by, and somewhat paternal to, Whitney. However, this early bemusement changed to raging opposition with Whitney's development of the Soil Survey. This controversy hinged on Whitney's inexplicable dismissal of the importance of soil chemical analyses to agriculture, and his belief that soil physical properties were the primary way in which soils affected vegetation. Hans Jenny, in his biography of Hilgard, paints a detailed portrait of this conflict, one in which Hilgard, by his own choice isolated in California, could only agitate against but could not change. The new soil survey, and scientific investigations associated with it, centered squarely on soil physical properties and analyses. Hilgard rose to the defense of his colleague, F.H. King, who had been Head of the Soil Management Investigations of the soil survey under Whitney. King was forced out of this position by Whitney, and three of his manscripts were censored by Whitney, forcing King to publish them privately. Hilgard rallied to King's defense in *Science*, and positive support for his defense came from such prominent scientists as T.C. Chamberlin.

Whitney's ascent corresponded with Hilgard's waning but, in typical fashion, highly productive years on the Berkeley faculty. Hilgard was obligated to retire in 1906, the first 'victim' of the university's emeritus policy. He continued to write. His book *Soils* was published 1906; and he coauthored a book on *Agriculture for Schools of the Pacific Slope*, and published numerous articles and comments.

The late nineteenth and early twentieth century was also a period for Hilgard to receive the awards and accolades that his work warranted. He was only the second graduate of Heidelberg to receive the 'Golden Degree,' commemorating the fiftieth anniversary of his dissertation. He received honorary doctorate degrees from Columbia, Michigan, Mississippi, and Berkeley. In regards to the latter award at Berkeley, received at a relatively advanced age in 1914, Hilgard remarked that he lived to see this award because "my persistence in living beyond the scriptural years carried the day." When Hilgard died, in early 1916, of pneumonia, the memorial accolades from the University of California and the National Academy of Sciences give the modern reader a keen appreciation of his stature, and the regard and esteem in which he was held.

Historians of science frequently note the fleeting nature of scientific stature and prominence in a human enterprise that unavoidably honors and discusses only the most recent publications, concepts, and trends. Even on the Berkeley campus, and from within Hilgard Hall where I write this biography, it is rare to find modern scholars who are familiar with Hilgard's life and contributions. Yet, in soil science, and in pedology in particular, his ground-breaking work and remarkable concepts resonate in much of what we do and think, in most cases without citation. Even courses still taught on the Berkeley campus originate from those first taught by Hilgard in the late nineteenth century, again without recognition. So, even at the beginning of the twenty-first century, his intellectual influence still persists, albeit in muted tones.

But Hilgard's contemporaries, possibly aware of the transient nature of academic fame, also chose to honor his contributions in ways of great permanence. To this day, Hilgard Avenues are found in Berkeley and Los Angeles, *Hilgardia* exists as a publication of the University of California, and Mount Hilgard, a towering massif of 13 361 ft, is found in a 'hall' of Sierra Nevada peaks named in honor of the great natural scientists of the nineteenth century. It is enlightening to read of the naming of Mount Hilgard as described by the surveyor who explored that region: "Mount Hilgard from the west is a stiking mass . . . It was named at the suggestion of an admiring former pupil of Professor Hilgard, Mr. Ernest C. Bonner, who accompanied me on one of my outings." It is instructive that a natural monument of such magnitude be chosen by a member of the generation that followed and was inspired by Hilgard. His work will surprise and stimulate those of the present generation who wish to step back in time and walk again with one of the giants of the natural sciences.

Further Reading

Amundson R and Yaalon D (1995) Hilgard EW and John Wesley Powell: efforts for a joint agricultural and geological survey. *Soil Science Society of America Journal* 59: 4–13.

Hilgard EW (1860) *Report on the Geology and Agriculture of the State of Mississippi*. Jackson, MS: E Barlesdale.

Hilgard EW (1884) General discussion of the cotton production of the United States: embracing the cotton-oil industry, methods and utility of soil investigation, and tables of cotton fiber measurements. In: Hilgard EW (ed.) (special agent in charge) *Report on the Cotton Production in the United States; Also Embracing Agricultural and Physico-geographic Descriptions of the Several Cotton States and of California*, vol. 5, pp. 7–93. Washington, DC: US Census Office.

Hilgard EW (1892) A report on the relations of soil to climate. *US Department of Agriculture Weather Bureau Bulletin* 3: 1–59.

Jenny H (1961) *E.W. Hilgard and the Birth of Modern Soil Science*. Pisa, Italy: Collana della Rivista agrochimica.

Slate F (1919) Biographical memoir of Eugene Woldemar Hilgard. *National Academy of Sciences Biographical Memoirs* 9: 95–155.

Wheeler BI, Wickson EJ, Loughridge RH, and Smith EA (1916) In memoriam: Eugene Woldemar Hilgard. *University of California Chronicle* 18: 1–50.

HOOGHOUDT, SYMEN BAREND

P A C Raats, Wageningen University, Wageningen, The Netherlands
R R van der Ploeg, University of Hannover, Hannover, Germany

Introduction

In the absence of dikes against sea and river water, more than half of The Netherlands would be flooded. Moreover, The Netherlands has a humid climate with, on average, precipitation exceeding evapotranspiration during autumn, winter, and early spring. Historically, crop yield records show a strong correlation between wet autumn and winter seasons and later low crop yields. However, from about 1850, the ability to get rid of excess water was the determining factor in crop yield. In the first half of the nineteenth century, much effort was put into modernizing the design, maintenance, and management of waterways. Replacement of windmills by steam-driven pumping stations started in 1845 and made removal of water independent of the whims of the weather. At the field scale, originally the removal of water was by a ridge-and-furrow system or a dense network of trenches. Already early in the nineteenth century, trenches filled with brushwood covered with soil were used incidentally, but such drains were found to decay rapidly. Earthenware tiles were introduced in England around 1845, and the first experiments in The Netherlands with these ‘drying pipes’ were done in 1851. Gradually, ditches were replaced by tiles of 0.05 m inside diameter, with a slope of 0.01–0.02 m m^{-1}, and draining into the remaining main ditches spaced 100–300 m apart, the larger spacings always being used in new polders. The choice of drain depth and distance depended on the nature of the soil and the drainage requirement. Leading the introduction of tile drainage, especially on arable land, was the province of Groningen, in the northeast, followed by the province of Zeeland, in the southwest, and the polders being reclaimed from the IJsselmeer. When, around 1960, tube drains succeeded tile drains, the practice also spread to other parts of the country. The drains not only serve to remove water, but in late spring and summer may also supply water from the main ditches to the fields by subirrigation.

In the 1850s, drainage experiments were conducted by the French engineer S.C. Delacroix. In the 1860s this work was noted by the Danish engineer D. Hannemann, who in turn inspired Ludvig August Colding (1815–88) to derive the ellipse equation. This Dupuit-type equation describes the shape of the water table for steady flow Q, resulting from uniform input s at the water table, to tile drains spaced l apart and resting on an impermeable base (Figure 1):

$$Q = ls = \frac{4kH_0^2}{l} \qquad [1]$$

with k representing the hydraulic conductivity of the soil and H_0 the height of the water table above the impermeable base at the midpoint between the drains. The ellipse equation was widely used in Denmark and also became known in other European countries. In The Netherlands it was used only sporadically, but nevertheless became the point of departure for S.B. Hooghoudt’s drainage research in the early 1930s.

From Physical Chemistry to First Contacts with Drainage Theory and Practice

Symen Barend Hooghoudt (Figure 2) was born on August 31, 1901 at Slochteren, in the province of Groningen, in the Netherlands. He studied physical chemistry at the State University of Groningen, receiving a BSc degree in 1922, an MSc degree in 1926, and a doctoral degree in 1928. His doctoral thesis, supervised by Frans Maurits Jaeger (1877–1945), dealt with a method for the accurate measurement of the Becquerel effect. This photovoltaic effect is the basic physical process through which a solar cell converts sunlight into electricity.

Hooghoudt joined the Soil Science Institute at Groningen on March 16, 1929 and remained with it till his death on August 30, 1953. In the late 1920s, the Institute’s Director, David Jacobus Hissink

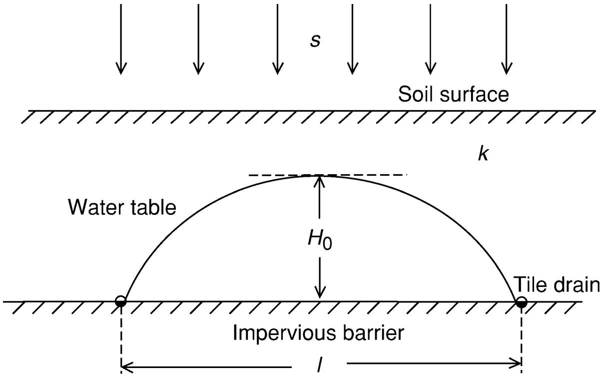


Figure 1 Flow to tile drains, as considered by L.A. Colding in 1872.

Figure 2 Symen Barend Hooghoudt (1901–53). (Reproduced with permission from Bruin P and Visser WC (1953) Het werk van Dr S.B. Hooghoudt in dienst van het landbouwkundig onderzoek. *Landbouwkundig Tijdschrift* 65: 650–657, © KLV Royal Netherlands Society for Agricultural Sciences.)

(1874–1956), was heavily involved in discussions within the International Society of Soil Science (ISSS) on the measurement of soil pH and its use as a diagnostic tool. The chinhydron electrode was the best method available, but unfortunately it could not be used in soils high in manganese and pH >8.5. Hooghoudt's first assignment was to study the antimony electrode as an alternative for soils with pH >8.5. Soon after this study was completed successfully, the glass electrode was introduced, allowing simple and rapid measurement up to pH 12–13, and as a result the antimony electrode was widely considered obsolete almost immediately.

In 1930–31 Hooghoudt was involved with Hissink in a first study of some physical properties in a wide range of Dutch soils, namely the permeability for water, the dry bulk density, and the air capacity, using methods devised by J. Kopecky early in the twentieth century. The prime motivation was the need for guidance in developing drainage systems for the planned IJsselmeerpolders, in particular the pilot polder Andijk and the first full-scale polder Wieringermeer. For reference, subsidence and prognosis of temporal changes in soil properties were studied in a series of polders reclaimed in 1550, 1844, 1924, and 1925. Several fields devoted to studies of drainage or subirrigation and an area intended to be used for a complex of sports and playing fields were also included in the study. The practical context exemplified by these field sites remained a hallmark of Hooghoudt's research.

The soil permeability was expressed in terms of the D-value, defined as the amount of water in meters per day flowing under a ponding depth of 0.04 m vertically through a sample of 0.07 m freely into the air. Despite its operational nature, the D-values enabled Hissink and Hooghoudt to delineate important aspects of the flow of water in Dutch soils. It was shown that for sandy soils without structure the particle-size distribution can be used to estimate the D-value, but that large structural pores dominate the conduction of water in all other soils. In soils with such structural pores, the D-values measured with ring samples of $3850\,\mathrm{mm}^2$ cross section were found to be extremely variable, even at very small distances.

In December 1931, Hooghoudt went on a study tour to Breslau, Prague, and Zurich. At Breslau, F. Zunker studied the dependence of various physical properties upon the specific surface derived from the particle-size distribution. In his trip report, Hooghoudt emphasized that Zunker's methods could only be used for sandy and very light loamy soils. At Prague, Hooghoudt learned about many practical aspects of soil and water management. But he found little of direct use, mainly because the structure of the soils differed too much from the generally rather young Dutch soils with comparable distributions of particle sizes, and also because some of the concepts being used were somewhat esoteric. The visit strengthened his belief that Dutch soils reclaimed from the sea have unique properties and that field methods were needed to determine soil properties relevant to the drainage process. The visit to E. Diserens at Zurich was by far the most fruitful. It brought Hooghoudt in contact with an analysis of drainage problems based on what he called the law of Dupuit–Darcy, and the associated laboratory and field methods for determining the required soil physical properties.

In July 1932, the 6th Commission of the ISSS held a meeting at Groningen on water management, with emphasis on drainage and reclamation of land from the sea. Together, the ISSS contacts, the impressions from the European study tour, the first study of soil physical properties with Hissink, and the practical needs of water management in The Netherlands shaped Hooghoudt's research plans for the next 10 years.

Physical Characterization of Soils in the Laboratory

From the very beginning, attempts were made to relate the physical properties of soils to the composition of the solid phase and the volume fractions of the solid, aqueous, and gaseous phases. The composition of the solid phase was given in terms of the contents of humus and calcium carbonate, and the content and size distribution of particles other than humus and calcium carbonate. At Groningen, the size distribution of particles was originally determined using either the Atterberg sedimentation cylinder method for the range 2–30 μm or the Kopecky multiple-tube elutriation method for the range 16–100 μm, complemented with sieving of the coarse fraction to 2000 μm. Hooghoudt made very detailed evaluations of these and other methods. The 323-page report of this methodological study includes a 59-page review of validity, limitations, and extensions of the Stokes law for the rate of fall of solid particles in a fluid. This report was submitted in August 1944. As for many scientists, for Hooghoudt scholarly work of this kind was a way to cope with the stagnation caused by the war-time conditions. In 1946 this study resulted in a switch to the combined sieve–pipette method for routine determinations, the pipette method being used for the fraction less than 35 μm and the sieve method for the fraction greater than 35 μm.

Following F. Zunker, Hooghoudt introduced the specific surface, U, which is defined as the ratio of the surface area of a given mass of soil divided by the surface area of this same mass of soil in the form of spheres having a reference diameter $\sigma_{\mathrm{r}} = 1\,\mathrm{cm}$. He showed that if $w = f(\sigma)$, where w is the weight fraction of the total amount of particles smaller than a specific particle diameter σ, then the specific surface U is given by:

$$U = \sigma_r \int_{w=0}^{w=1} \sigma^{-1} \mathrm{d}f = \sigma_{\mathrm{r}} \int_{\sigma=\sigma_1}^{\sigma=\sigma_2} \sigma^{-1} (\mathrm{d}f/\mathrm{d}\sigma)\mathrm{d}\sigma \qquad [2]$$

A hypothetical soil with uniform particle diameter $\sigma_{\mathrm{eff}} = \sigma_{\mathrm{r}}/U$ has the same specific surface U as the actual soil. Hooghoudt refers to σ_{eff} as the effective particle diameter of the actual soil. The specific surface area U and effective particle size σ_{eff} derived from it are particularly relevant in the range 16–2000 μm. Clearly, U and σ_{eff} are proper measures of the coarseness or fineness of a soil and σ_{eff} can be regarded as a characteristic microscopic length scale of a soil, comparable with that used in the Miller–Miller scaling theory of modern soil physics. Consistent with this, in expressions used by Hooghoudt, the hydraulic conductivity is inversely proportional to U^2 and the height of capillary rise is proportional to U.

From the very beginning, Hooghoudt distinguished two broad classes of soils: a class of soils without structure, and a class of soils with structure arising from aggregation, cracking, and perforation. For the class of soils without structure, the physical properties can be determined in the laboratory on disturbed samples and the results can be extrapolated to the field situation, using formulas involving the porosity, volume of entrapped air, and temperature. For the class of soils with structure in their natural condition, meaningful measurements of the hydraulic conductivity cannot be made in the laboratory, but must necessarily be made in the field. In fields without drainage systems, this can be done by the auger-hole method. In fields with drainage systems, the hydraulic conductivity can also be inferred from the relationship between the groundwater level and the flux into or out of drains and ditches.

The phase distribution of soils was characterized by the total porosity, considered as the sum of the drainable porosity and the water-filled porosity after drainage, and the air-filled porosity below the water table. Related to the dynamics of the water, Hooghoudt determined two physical parameters in the laboratory: the height of capillary rise and the hydraulic conductivity below the water table. To interpret data on height of capillary rise and hydraulic conductivity, he made use of rational formulas, showing the dependence upon the specific surface U and the porosity. By modern standards, the concepts used by Hooghoudt may seem simplistic, but they were perfectly suited for his goals.

Hooghoudt also paid some attention to rheological properties of soils, resulting in a 200-page publication on hardness and on rupture by various means of soils in the dried state. The aim was to get a basic understanding of the mechanical properties of dry clods and crusts. He used methods suitable for dried pastes borrowed from research related to the pottery and porcelain industry. He focused on the influence of particle-size distribution and the composition of the adsorption complex on the mechanical properties.

In the context of a large-scale study of irreversible drying of peat soils, Hooghoudt studied the causes of the phenomenon and developed a method to express the degree of irreversible drying by a single number.

Auger-Hole Method for Determining Hydraulic Conductivity in the Field

Observations of the rate of rise of the water level in an auger-hole following lowering of the water table by pumping can be used to calculate the hydraulic conductivity. E. Diserens and J. Donat pioneered this method. Following his discussions with Diserens in December 1931, Hooghoudt greatly improved the underlying theory, arriving for homogeneous soils at a formula expressing the rate of rise of the water level as the sum of a term due to horizontal flow through the wall of the hole and a term due to vertical flow through the bottom of the hole. He also derived corresponding formulas allowing for soil heterogeneity, either in the form of layering or continuous variation with depth. The formulas were verified by laboratory tests in a 62.5-cm-diameter drum with a perforated copper tube placed at the center. In the experiments the diameter of the tube, the depth of the bottom of the tube below the undisturbed water table, the initial lowering of the water in the tube, and the location of the impermeable layer were varied. Both homogeneous and layered soils were used in the tests. To study the possible influence of the small diameter of the drum, tests were also performed in a soil bin 10 m long, 2 m wide, and 2 m deep.

By necessity the tests of the theory in the small soil drum and large soil bin were done with sandy soils. However, Hooghoudt also developed the equipment needed for measurements in the field, allowing measurements to be made not only in sandy soils, but also in structured soils. With this equipment, measurements were made in the Wieringermeerpolder in a drained field, allowing comparison with hydraulic conductivity inferred from the relationship between groundwater levels and drain fluxes. The agreement between the hydraulic conductivities determined with the two methods was quite satisfactory, and Hooghoudt proceeded to use the auger-hole method widely and used the results to give advice with regard to design of drainage systems. In his extensive 1936 report on the auger-hole method, he gave as examples data for a proposed airfield near Leeuwarden and for the Rietwijkeroorderpolder near Amsterdam, intended to be developed as a recreational park.

The theory underlying the auger-hole method was developed further after World War II. At Groningen, L.F. Ernst in 1950 used J.J. van Deemter's numerical relaxation method to obtain a new formula to calculate the hydraulic conductivity from auger-hole measurements. In the USA, Don Kirkham and his associates made important contributions from 1948 onward. Based on their very general analytical theory for auger-hole seepage, C. Boast and D. Kirkham in 1971 concluded:

> Hooghoudt's formula should be used only for a restricted range of geometries. Ernst's formula or graph, or table we give should be used otherwise.

Flow of Water in Shallow Soils Without or With Structure in Their Natural Condition

In a 1937 publication, Hooghoudt developed a theory for flow to ditches and drains in shallow soils. Building upon the hydraulic approach of Colding and his followers, he assumed that the vertical distribution of the pressure is hydrostatic at any point in the field, and that as a consequence the slope of the water table can be regarded as the driving force for the flow in the horizontal direction. For structured soils, he ignored the contribution from horizontal flow in a capillary fringe. For a steady flow Q resulting from uniform input s at the water table, to ditches spaced l apart and reaching to the impermeable base, and with a water table everywhere below the soil surface, Hooghoudt derived the following expression for the flux Q (Figure 3):

$$Q = ls = \frac{4k\left(H_0^2 - h_0^2\right)}{l} = k\frac{8h_0 + 4m_0}{l}m_0 \qquad [3]$$

where H_0 is the height of the water table above the impermeable base at the midpoint between the drains, h_0 is water level in the ditch, and $m_0 = H_0 h_0$. If the hydraulic conductivity k is known, say from measurements with the auger-hole method, then Eqn [3] can be used to calculate the desired drainage spacing l for given s, H_0, and h_0. Alternatively, if the hydraulic conductivity k is not known, it can be calculated if l, s, H_0, and h_0 for a drainage system are known.

To verify Eqn [3], Hooghoudt used a soil bin 10 m long, 2 m wide, and 2 m deep, filled with river sand. He found that the estimates of hydraulic conductivity on the basis of Eqn [3] agreed very well with laboratory data, provided the effect of flow in the capillary fringe of thickness h_c was included by replacing H_0, and h_0 by $H_0 + h_c$ and $h_0 + h_c$.

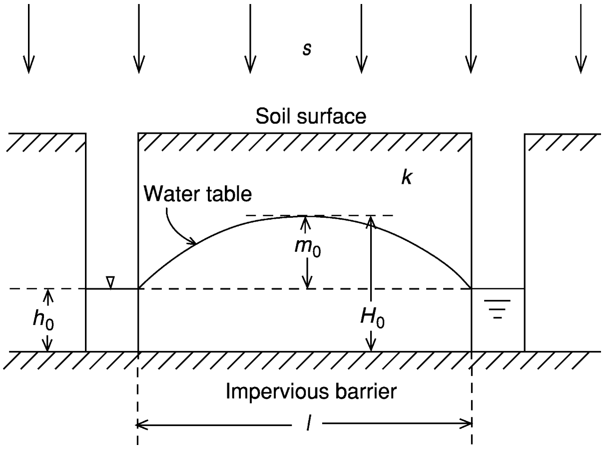


Figure 3 Flow to ditches, as considered by Hooghoudt in 1937.

Still for homogeneous soils, Hooghoudt considered various complicating factors:

- Situations with the input s exceeding the value for which the water table at the midpoint between the ditches or drains reaches the soil surface;
- Situations with the drain or the bottom of the ditch being some distance above the impermeable base. Tentatively these were handled by increasing l in the denominator of Eqn [3] by $r(r+b)/(2h_0)$ where b is the width of the bottom of the ditch ($b=0$ for a drain) and r is the depth of the impermeable layer below the bottom of the ditch or drain;
- Situations with flow not only to parallel drains and small ditches but also toward a larger ditch that intercepts the flow from the parallel drains or small ditches; for this case he also used the large soil bin to test the results.

Hooghoudt also extended the hydraulic approach to heterogeneous soils, considering in detail cases with jump discontinuities of the hydraulic conductivity at interfaces between horizontal, homogeneous layers, and cases with gradual changes of the hydraulic conductivity with vertical position. In effect he replaced, at any point between the drains, the hydraulic conductivity k by an apparent hydraulic conductivity k_{av}, being defined as the vertically averaged conductivity between the impermeable base and the location of the water table. Clearly, even with the heterogeneity being restricted to the vertical direction, k_{av} is inherently dependent on the location of the water table. Nevertheless, the results of the analysis can still be written in the form of Eqn [3], provided k is replaced by an effective conductivity k_{eff}, which is a function not only of the vertical distribution of the real conductivity, but also of H_0 and h_0. For example if the real conductivity k increases linearly with height y according to $k=k_0+ay$, where a is a constant, then the effective conductivity k_{eff} in an expression of the form of Eqn [3] will be given by:

$$k_{\mathrm{eff}} = k_0 + \frac{\frac{1}{3}a\left(H_0^3 - h_0^3\right)}{H_0^2 - h_0^2}. \qquad [4]$$

Note that if $a=0$ then k_{eff} reduces to k_0.

The Hooghoudt Drainage Equation

In a 1940 publication, Hooghoudt removed the restriction of the impermeable base being relatively close to the drain or the bottom of the ditch. To this end he introduced the device of partitioning the flow region in two parts, one part away from the drain or ditch in which the flow is approximately horizontal and another part close to the drain or the ditch in which the flow is radial. The result was cast in a form similar to Eqn [3], with the radial resistance being accounted for by replacing h_0 by an equivalent thickness h_{eq} of the water-conducting layer below the drain, and allowing for the hydraulic conductivity in the layers above and below the drain having different values of, respectively, k_1 and k_2 (**Figure 4**):

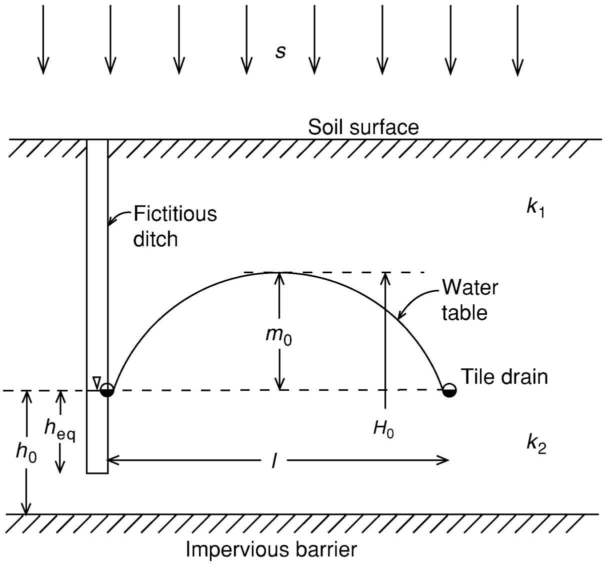


Figure 4 Flow to tile drains, as considered in the Hooghoudt drainage equation of 1940.

$$Q = ls = \frac{8k_2h_{\mathrm{eq}} + 4k_1m_0}{l}m_0 \qquad [5]$$

with the equivalent thickness h_{eq} being a known function of drain spacing l, the real thickness of the water-conducting layer below the drain h_0, and the diameter of the drain. Eqn [5] is usually referred to as the Hooghoudt drainage equation.

Applications of Drainage Theory

The planning of reclamation of the IJsselmeerpolders stimulated a rapid development of the study of drainage of soils for agricultural purposes, with theory and practice going hand in hand. D.J. Hissink and his staff started activities in the pilot polder Andijk in the late 1920s and made extensive studies in the Wieringermeerpolder in the early 1930s. Hooghoudt was responsible for the scientific basis of the design of the drainage systems. In the same period, smaller polders were also studied. For example, for the Linthorst Homanpolder in the north of the province of Groningen a drainage plan was developed, and for the Nijkspolder in the southeast of the province of Friesland a study was made of the expected increase in seepage into the polder as a result of an increase in the size of this polder and a lowering of the water levels in the open water within the polder. On an even smaller scale, numerous drainage recommendations were given for individual farmers' fields.

Three reports concerning sports fields were submitted in 1938. In the post World War II period 1947–51, 15 more reports followed. Recommendations included design of drainage systems and alterations of the soil profile to improve playability by changes in layering and by supplementing clay soils with sand. A comparison of the 1939 and 1953 guidelines for the design of soccer fields illustrates the progress.

Hooghoudt and his colleagues J.D. de Jong and A.J. Wiggers assisted the Laboratory of Soil Mechanics at Delft in a government assignment to make recommendations regarding suitability and possible improvements of sites for civilian and military airfields. Two reports concerning the military airfield Valkenburg in the province of South Holland were presented in 1939. In the post World War II period 1945–48, about 20 more reports were presented, concerning nearly as many locations.

The Committee Forestry Plan Amsterdam prepared an ambitious plan for a recreational park at the southwestern edge of the city of Amsterdam. Hooghoudt was asked to design a drainage plan. In the period 1936–39, Hooghoudt submitted eight reports concerning the soils in the area, including measurements of the hydraulic conductivity by the auger-hole method, and recommendations for spacing and depth of drains. Based on these, 300 km of tiles were laid.

Hooghoudt was also responsible for the first regional hydrologic study. In the Veenkoloniën, the thickness and the hydraulic conductivity of the water-conducting layer were studied to a depth of 20 m. On the basis of these data, the functioning of canals and ditches was analyzed. It was the first study at this scale, allowing a prognosis of the consequences of a variety of measures such as eliminating or reducing the size of canals originally dug for the purpose of transport of peat being mined for fuel, eliminating superfluous ditches, and the breaking of poorly conducting layers inside and underneath remaining peat layers.

Water Management of the Plant Root Zone

All along, it was of course realized that generally the purpose of drainage and subirrigation is to promote a favorable environment for the growth and functioning of plant roots. In February 1936, Otto de Vries (1881–1948), son of the botanist Hugo de Vries (1848–1935), gave a lecture on 'Soil, water, and plant' at the first meeting of the Dutch section of the ISSS. Following his 22 years in agricultural chemistry in the Dutch Indies, in 1930 Otto de Vries became director of the Agricultural Experiment Station and Soil Testing Laboratory at Groningen and, in 1939, following the retirement of Hissink, director of the Soil Science Institute. In 1943 the Departments for Agricultural Research and for Agricultural Industries of the Central National Council for Applied Scientific Research in the Netherlands (TNO) were founded and Otto de Vries was appointed chairman of both. He left Groningen in 1944 and lived in The Hague until his death in 1948.

In 1942 a field experiment for studying the influence of controlled water table depth upon the growth and yield of crops was realized at Nieuw Beerta in the eastern part of the province of Groningen. The experiment was continued till autumn 1955. The detailed description and interpretation of the results in eight publications illustrate Hooghoudt's evolving interest in the agronomical effects of drainage. Of particular interest is the early demonstration that crop response to drainage is complicated, being mediated by the influence of the water table depth on soil structure, plant root development, and the fate of soil nitrogen.

In 1943 a national Working Committee for Evaporation Research was formed, and in 1946 the more broadly based Committee for Hydrological Research TNO. Under the leadership of Hooghoudt, plans were made for a study of the evaporation from a polder, comparing three methods: (1) determining the water budget of the entire polder, (2) direct measurement of evaporation on the basis of a calculation of vertical vapor transport from measured vertical humidity and wind velocity profiles by the Royal Dutch Meteorological Institute, and (3) determining the soil-profile water balance using a lysimeter and soil-moisture sampling. For the latter purpose, the use of tensiometers and thermal conductivity probes was initially considered, but later abandoned for various reasons. The spatial variability of the soil moisture was recognized as a major obstacle. Hooghoudt played a leading role in the choice of the location and the detailed planning of this experiment. Perhaps not surprising, from a large number of possibilities the Rottegatspolder near Groningen was selected. Data collection started August 1, 1947 and terminated February 1, 1971, thus long after the agrohydrologic research had moved in 1956 from Groningen to Wageningen.

Mentor of Young Talents: The Theorists Van Deemter and Ernst

After World War II, Hooghoudt concentrated on the application of drainage theory and the exploration of its agronomical implications. Eventually he coached a large team working throughout the country. Further development and experimental verification of the theory were the responsibility of, respectively, J.J. van Deemter and L.F. Ernst.

Jan Josef van Deemter, after earning in 1945 an MSc degree in physics from the University of Groningen, was employed in the Agrohydrology group from June 1946 till some time in 1947, when he joined the Royal/Dutch Shell Laboratory at Amsterdam. On the basis of the results of his research started in the brief period at Groningen, in 1950 he was awarded a doctoral degree in applied mathematics at the University of Amsterdam. In his thesis, he adapted the hodograph method to give exact solutions of the problems of drainage and infiltration for systems of parallel, equidistant drains or ditches in soils with homogeneous hydraulic conductivity to infinite depth, including the effects of rainfall, evaporation, and upward or downward flow. He also used the so-called relaxation method of Southwell, which can be applied to all two-dimensional steady-state problems, as well as for soils with heterogeneous hydraulic conductivity.

After van Deemter left Groningen, he was succeeded by Lodewijk Ferdinand Ernst, who improved the auger-hole method and generalized the drainage formulas. Ernst became involved in several large-scale field studies and continued research in the style of Hooghoudt after he moved to Wageningen in 1956. In 1962, Ernst was awarded a doctoral degree in applied mathematics at the University of Utrecht on the basis of a thesis about groundwater flow in the saturated zone in the presence of parallel open conduits.

Hooghoudt's Legacy

In the autumn of 1952, there was a proposal to move the Agricultural Experiment Station and Institute for Soil Research TNO partly or entirely to a more central location at Wageningen. Hooghoudt was strongly opposed to such a move and pointed out that in his experience it was quite possible to lead joint research projects throughout the country from Groningen. At a staff meeting on December 9, 1952, he illustrated this with a description of recent and current agrohydrologic research by him and his team, emphasizing the great advantage of having agronomic and hydrologic expertise under one roof. Theory, experiment, and practical application were perfectly balanced in the research program of Hooghoudt.

Hooghoudt died at Groningen on August 30, 1953. At a memorial meeting on October 1, 1953, the director, P. Bruin, gave a detailed survey of his work. Three years later the agrohydrogic program was moved to Wageningen to become the core of the newly formed Institute for Land and Water Management Research (ICW). At the ICW, Wageningen University, and the International Institute for Land Reclamation and Improvement (ILRI; founded in 1955), Hooghoudt's ideas flourished and were promoted worldwide. The ILRI International Course on Land Drainage has, since its first session in 1962, drawn more than 1000 drainage professionals from all over the world. Dutch-born PhD students at Iowa State University and Cornell University have facilitated the export of some of Hooghoudt's ideas published mainly in Dutch to the USA and the English-speaking world in general. This was consolidated by the 1957 monograph *Drainage of Agricultural Lands*, published by the American Society of Agronomy, as well as in the later, updated editions of 1974 and 1999.

See also: **Drainage, Surface and Subsurface**

Further Reading

Bear FE (ed.) (1952) *Soil Science* 74: 1–101.

Bruin P and Visser WC (1953) Het werk van Dr S.B. Hooghoudt in dienst van het landbouwkundig onderzoek. *Landbouwkundig Tijdschrift* 65: 650–657.

Childs EC (1969) *An Introduction to the Physical Basis of Soil Water Phenomena,* Chapt. 16. London, UK: John Wiley.

De Vries JJ (1982) *Anderhalve Eeuw Hydrologisch Onderzoek in Nederland*. Overzicht van de ontwikkeling van de wetenschappelijke kennis van het water in Nederland tussen 1830 en 1980, beschouwd vanuit een geohydrologische gezichtshoek. Amsterdam, The Netherlands: Rodopi.

Harmsen K (1990) *Het Instituut voor Bodemvruchtbaarheid 1890–1990*. Haren (Gr), The Netherlands: Knoop.

Hissink DJ and Hooghoudt SB (1931) Bijdrage tot de kennis van eenige natuurkundige grootheden van den grond. Eerste mededeeling [Contribution to the knowledge of some physical parameters of soils. First commnunication]. *Reports of Agricultural Research* 37B: 101–190. The Hague, The Netherlands: General Government Printing and Publishing Office.

Hooghoudt SB (1934–1950) Bijdragen tot de kennis van van eenige natuurkundige grootheden van den grond, 2–10 [Contribution to the knowledge of some physical parameters of soils, 2–10]. *Reports of Agricultural Research* 40B: 215–345 (1934); 41(1)B: 589–615 (1935); 42(13)B: 449–541 (1936); 43(1)B: 1–11 (1937);

43(13)B: 461–676 (1937); 46(14)B: 515–707 (1940); 49(16)A: 721–913 (1943); 50(13)A: 671–993 (1945); 56(3): 1–209 (1950). The Hague, The Netherlands: General Government Printing and Publishing Office.

Luthin JJ (1957) *Drainage of Agricultural Lands*. Madison, WI: American Society of Agronomy.

Van Deemter JJ (1950) Bijdragen tot de kennis van enige natuurkundige grootheden van de grond, no. 11. Theoretical and numerical treatment of flow problems connected with drainage and irrigation (in Dutch). *Reports of Agricultural Research* 56.7: I–VIII, 1–67. The Hague, The Netherlands: General Government Printing and Publishing Office.

Van der Loeff JA (1954) *Publications and reports of Dr. S.B. Hooghoudt*[†], *Principal Scientific Officer with the Agricultural Research Station and Soil Science Institute T.N.O. Groningen* (in Dutch). Literatuurlijst nr. 12. Centrum voor Landbouwdocumentatie. The Hague, The Netherlands: Ministerie van Landbouw, Visserij en Voedselvoorziening.

Van der Ploeg RR, Kirkham MB, and Marquardt M (1999) The Colding equation for soil drainage: its origin, evolution, and use. *Soil Science Society of America Journal* 63: 33–39.

Van der Ploeg RR, Marquardt M, and Kirkham D (1999) On the history of the ellipse equation for soil drainage. *Soil Science Society of America Journal* 61: 1604–1606.

Van Schilfgaarde J, Kirkham D, and Frevert RK (1956) Physical and mathematical theories of tile and ditch drainage and their usefulness in design. Research Bulletin 436. Ames, IA: Iowa Agricultural Experimental Station.

Zunker F (1930) Das Verhalten des Bodens zum Wasser. In: Blanck E (ed.) *Handbuch der Bodenlehre*, vol. 6, pp. 66–220, *Die physikalische Beschaffenheit des Bodens*. Berlin, Germany: Springer-Verlag.

HUMIFICATION

T C Balser, University of Wisconsin–Madison, Madison, WI, USA

Introduction

Soil organic matter represents the largest dynamic, terrestrial carbon pool. While many aspects of the processes controlling its dynamics are understood, the mechanisms leading to the accumulation and storage of carbon in soil have not been substantially elucidated. The quantity of carbon in soil is greater than that in vegetation and the atmosphere combined (Figure 1), and a large part of that carbon is in a stable form, either physically protected or chemically recalcitrant (humified). Physically protected carbon is that associated with soil mineral surfaces. The process and degree of physical protection are topics that have been widely studied in order to quantify and predict carbon sequestration in soil. Soils with differing mineralogy will protect and sequester carbon to varying degrees (Figure 2). Humification (the alteration of biologically derived carbon to chemically complex forms) also represents a critical process driving carbon sequestration. Because humic substances (HS) in soil are highly recalcitrant, with turnover times ranging to thousands of years, humified carbon is sequestered carbon. While carbon sequestration has been the subject of attention recently due to changes in global climate and atmospheric chemistry, it has been known for more than a hundred years that soil carbon (particularly HS) is essential to soil fertility and productivity. HS are important regulators of trace metal concentrations in soil solution and may act as plant growth promoters. In addition, HS are critical determinants of soil properties such as water-holding capacity, structure and aggregation, and nutrient retention and exchange.

Soil microorganisms are dominantly responsible for the formation and turnover of HS in soil. Yet models

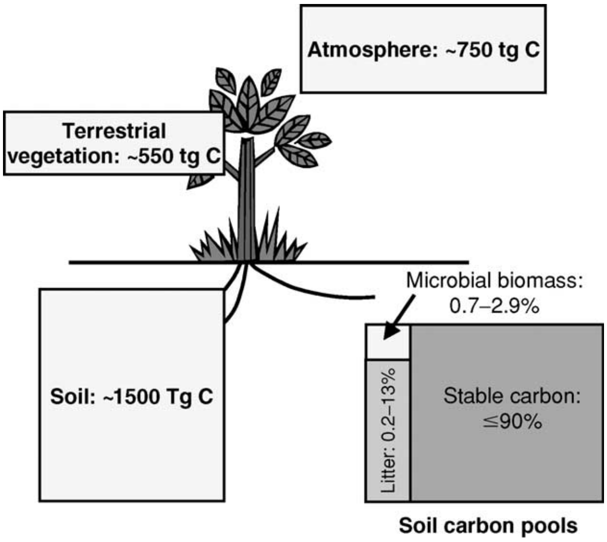


Figure 1 Global carbon pools. Soil carbon content is greater than that in terrestrial vegetation and the atmosphere combined. Up to 90% of soil carbon is in stable, humified form. Tg, teragrams.

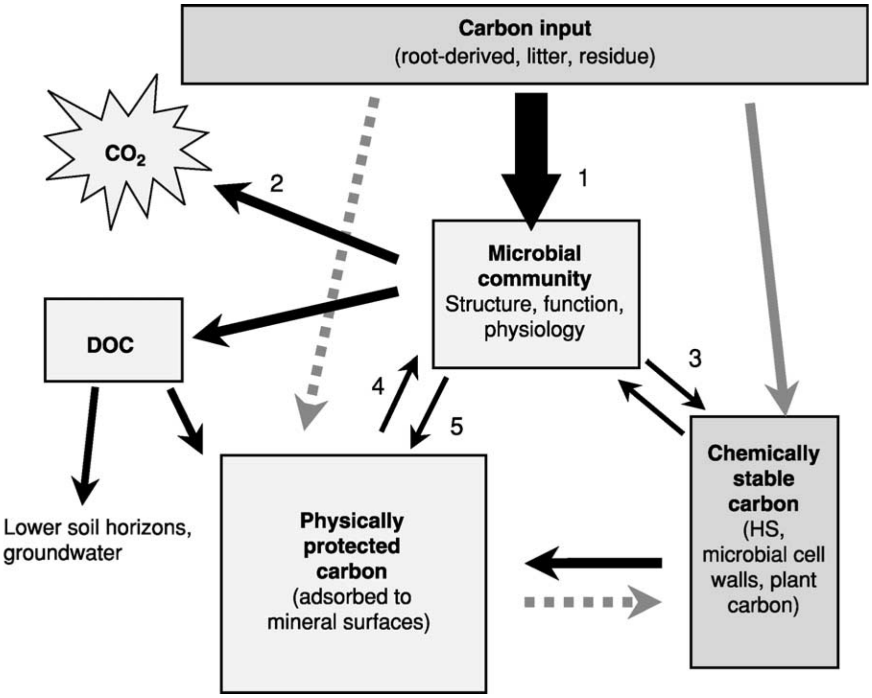


Figure 2 Soil carbon storage showing microbial community control. Community physiology and adaptation to incoming carbon quality determines partitioning between loss as CO_2 or dissolved organic carbon (DOC), and chemically stable or physically protected forms (1–3). Gross scale community composition determines contribution of cell walls to stable C as senescent biomass. Soil mineralogy can alter all fluxes, especially between the physically protected pool and microbial community (4–5). HS, humic substances.

and field studies of carbon storage have tended to concentrate on the physical factors controlling the size of carbon pools, rather than on microbiological controls over carbon transformation in soil. Intuition suggests microorganisms play a critical role in determining the loss or storage of soil carbon, yet gaps remain in our understanding. The microbial role in promoting carbon loss from soil is obvious: heterotrophic metabolism results in production of CO_2. Even so, we do not fully understand the controls on heterotrophic metabolism in the soil. Microbial physiology can tell us how different chemical constituents might be utilized for biomass and energy, but will not elucidate the long-term implications of metabolism and physiology for carbon storage and humification. Similarly, typical studies of humification show changes in chemical structure, but we can only infer process from such studies. Studies of organic-matter chemistry cannot elucidate the mechanistic relationships that determine carbon partitioning between microbial biomass, CO_2, and HS.

There is a difference in perspective about the role of microorganisms in humification that arises from chemical versus microbiological studies of HS genesis and characterization. Chemical studies are designed primarily to elucidate structural details of soil carbon 'pools' and infer the process of HS genesis. This provides critical information on the behavior of HS in soil, and their role in soil fertility, water-holding capacity, and structure. In contrast, a microbiological perspective addresses the question of why humification happens, as well as where, and how rapidly. A dynamic, microbiologically based understanding of humification will be necessary in the context of current and future research in soil carbon cycling.

Historical Perspective

It has long been recognized that HS play a critical role in soil functioning. From the late 1700s, soil HS have been studied in the context of agriculture and soil fertility. The 1700s up to the 1970s saw major advances in our understanding of the nature and properties of soil HS, as wet chemical techniques gave way to increasingly sophisticated analytical equipment. In the 1970s, advances in organic geochemistry analytical techniques, coupled with interest from the petroleum industry, fueled research in humification and humus genesis and characterization beyond the agronomic realm. The 1980s saw growing interest in the role HS play in environmental and ecological arenas. HS are involved in disinfection by product formation in drinking water (e.g., organohalide formation), and in toxic-metal binding and bioavailability. The interaction between HS and soil, sediment, and water contaminants such as pesticides and polyaromatic hydrocarbons has received increasing attention. In addition, the role of humification as a mechanism of soil carbon sequestration has become an area of importance in light of current, global-scale changes in atmospheric carbon concentration.

With increasing emphasis on humification and carbon storage has come a renewed interest in understanding and quantifying the microbial role in carbon cycling. The first recognition that microorganisms play an important role in HS genesis came in the early 1900s, but the nature and importance of that role continue to be debated even today. Because of the long time scale necessary for humus genesis and its complex chemical nature, HS research has traditionally focused on isolation and characterization rather than process dynamics. The discipline has been dominated by organic chemists and geochemists rather than microbiologists, resulting in long-standing operational definitions of HS and an entire language used to describe the properties and nature of HS. Terms such as ‘humic acid,’ ‘fulvic acid,’ and ‘humin’ refer to classic soil-extraction procedures based on solubility in aqueous media, as a function of pH. While past chemical emphasis on the genesis and behavior of humic substances has contributed greatly to our understanding, there is a high degree of uncertainty regarding the microbial role played in humification. Yet, with the development of recent analytical techniques, there is ever-increasing evidence that microorganisms play an important role in the genesis and composition of HS in soil.

Chemical Perspective on HS Genesis

HS are by nature highly complex mixtures. Understanding how HS form, until recently, has required us in effect to blow up HS molecules, characterize the fragments, and then attempt to understand how they were put together in the first place (Figure 3). This is a bit like wrapping a watch in a handkerchief, smashing it with a sledge hammer, then handing the bits to someone and asking them not only to reconstruct the watch, but also to explain how the watch was formed in the first place! While this is the domain in which soil chemists and geochemists excel, this type of ‘bottom-up’ analysis is a challenge, to say the least. Ideally, an understanding of HS genesis could be gained from a ‘top-down’ approach, observing formation from biogenic materials. However, HS genesis takes place over a time scale of hundreds to thousands of years, rendering direct observation impractical. Thus, an understanding of HS genesis really does require an understanding of the chemical nature and properties of extant HS extracted from soil. Methods developed over the past several decades have led to ever-more sophisticated means of investigating HS chemistry, allowing more detail in structural inference. As methods have been developed and the knowledge base expanded, so too has the postulated role of microorganisms been reassessed (Table 1).

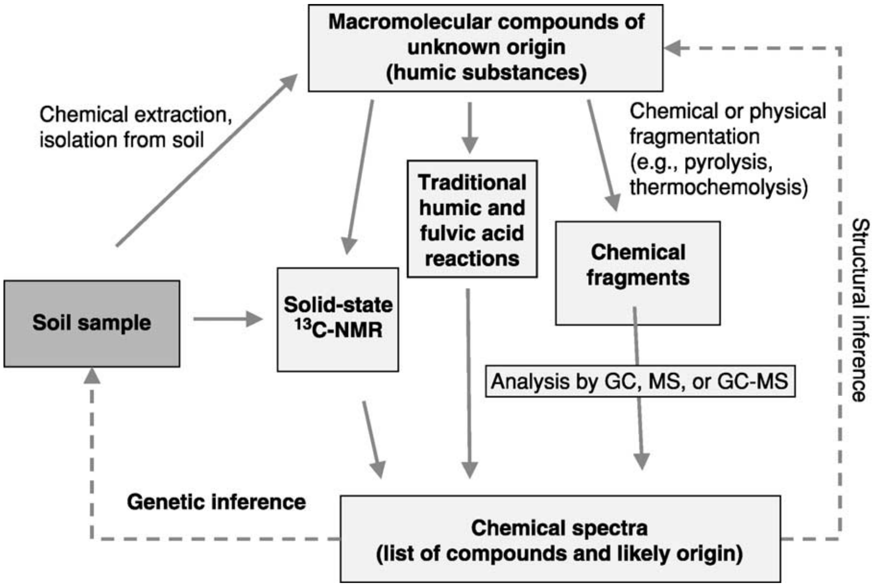


Figure 3 Flow diagram of deductive analytical techniques for investigating humic substances (HS) genesis and composition. GC, gas chromatography; MS, mass spectrometry; NMR, nuclear magnetic resonance.

Table 1 Comparison of traditional and newer methods for humic substances (HS) analysis

	Traditional degradative methods	*Nondegradative methods and combinations of newer techniques*
Method description	HS separated on the basis of chemical solubility (operational definitions rather than structural or compositional)	HS studied in unaltered form or in chemical or physical fractions
Structural inferences drawn	HS are macromolecular, highly aromatic	HS are macromolecular, highly aliphatic
Postulated microbial role	Metabolic action in early stages results in selective preservation of plant structural carbon, and abiotic condensation of smaller molecules	Large contribution of microbial cell and metabolic components into HS in addition to importance of metabolic role

Traditional (Degradative) Methods of HS Characterization

Common traditional methods for HS characterization have typically been 'degradative' approaches that involve the physical or chemical breakdown of macromolecular compounds into various subunits which are subsequently isolated and identified (Figure 3). Structural inferences are drawn based on the identities and yields of degradation products and the reaction processes responsible for their generation. This class of methods includes a range of extraction techniques (e.g., alkaline, organic, or neutral salt extractants) that are designed to solubilize and isolate chemical fractions from HS. For example, the humic acid fraction is not soluble in water under acid conditions, the fulvic acid fraction is soluble in aqueous solution under all pH values, and humin is neither soluble nor extractable in aqueous media. Traditional isolation methods begin by treating the sample with a strongly basic solution. Both humic and fulvic acids are coextracted in basic solution. The unextracted residue contains the humin. Upon acidification (with a strong acid such as HCl) of the alkaline extract, humic acids precipitate. Organic material remaining in solution is considered to be the fulvic acid fraction.

A drawback to this type of extraction is that it is extremely difficult to separate all nonhumic material from HS. Thus, not all the extracted material is HS. For example, long-chain aliphatic compounds may be closely associated with extracted HS. Also, the specific conditions chosen for extraction can produce chemical artifacts that may lead to false structural inferences. Strong alkaline extraction, in particular, can dissolve organic components from fresh plant material, incorporating them into humified organic matter. In addition, condensation reactions (e.g., between amino acids and reducing sugars or quinones) can occur in alkaline solution, resulting in humic materials created in the laboratory during extraction. These types of artifacts have led to the belief that humic substances are dominantly aromatic in structure (Table 1).

Nondegradative and Newer Degradative Techniques

Newer methods avoid the artifacts of traditional degradative techniques and indicate different structures for HS from those determined by traditional extraction techniques (Table 1). Nondegradative techniques are among the methods that may indicate the role of microorganisms in HS formation more reliably than past methodologies. Nondegradative approaches allow analysis of HS directly, usually without further chemical alteration. They include solid-state nuclear magnetic resonance (NMR), UV-visible and infrared spectrophotometry, electron spin resonance (ESR) spectrometry, X-ray analysis, electron microscopy, and electron diffraction.

A method that has received considerable attention is solid-state NMR spectroscopy. Liquid-state NMR has been used in chemical studies for a long time, but until relatively recently the complexity of HS and the presence of paramagnetic minerals (such as iron oxides) in soil has prevented the effective use of solid-state techniques. With the introduction of methodological refinements such as Fourier transform analysis and cross-polarization magic angle spinning (CP-MAS), researchers have been able to look at unaltered chemical mixtures in soil directly. A primary advantage of solid-state over liquid-state NMR is that most soil organic matter is insoluble and can only be analyzed in solid state. In addition, solid-state NMR is nondestructive – the sample is unaltered by solvent effects. Using NMR, it has been clearly shown that HS are not as aromatic in nature as previously thought. Studies of the composition of soil organic matter in relatively undisturbed soil profiles (forest floor samples, in particular) indicate a shift from plant-derived carbon near the top to carbon that has a microbial origin lower in the profile. Microbial components and metabolic products accumulate in HS as soil carbon becomes humified, indicating a significant microbial role in humification.

While nondegradative techniques have provided some of the first indications of the extent of microbial contribution to humification, there are still drawbacks to their exclusive use. For example, ^{13}C and ^{15}N-NMR require the mitigation of issues associated with solid samples, such as C-H and N-H dipolar interactions, chemical shift anisotropy, and interference by paramagnetic minerals (e.g., iron oxides) that can cause loss of sensitivity and line-broadening in the resulting spectra.

Advances in degradative techniques, and combinations of nondegradative and degradative techniques, have proven effective in allowing further understanding of HS chemistry and indicating the role of microorganisms in HS genesis. While new degradative techniques still involve the extraction of HS from soil, degradation and/or fragmentation, and analysis of the resultant product, changes in reaction conditions and method combinations reduce the incidence of chemical artifacts.

Thermal decomposition (pyrolysis) produces fragments of the main structural elements of HS. Pyrolysis with gas chromatography and mass spectrometry (GC-MS) has provided a significant new body of information about HS chemistry. A major advantage of pyrolysis is that it does not require solubility. During pyrolysis, the sample is heated rapidly in an inert environment. Thermal energy causes fragmentation of the HS macromolecules. The resulting volatile products are chromatographically separated (GC) and identified (MS). Alternatively, the volatile products can be ionized and identified using soft field-ionization mass spectrometry (FIMS). Pyrolysis GC-MS or FIMS yields data on the molecular composition of HS from the proportions of several classes of compounds in summed and averaged mass spectra. Furthermore, thermograms of volatilization give information on the strengths of chemical bonds within the organic macromolecules, or between organic materials and minerals. The thermal energy required to volatilize a compound indicates bond strength; and because bond strength is a proxy for resistance to microbial attack, pyrolysis-MS thermograms can indicate degree of humification. Drawbacks to pyrolytic techniques are that only a fraction of the sample is pyrolyzed, the high temperature can alter the structural 'building blocks' being studied, and secondary reactions between volatile products (such as rearrangement or polymerization) can occur. Structural inference must be drawn with caution. Despite these drawbacks, pyrolysis provides molecular-level detail about the original HS that is at least as representative of the original material as any other technique and requires very little sample. The molecular-level information provided complements the broader scale of information from NMR, and studies such as these have proven useful in indicating the role of microorganisms in humification.

The use of capillary GC (e.g., in pyrolysis GC-MS) is a powerful tool for resolution of compounds in complex mixtures, but the compounds must be of low molecular weight and low polarity. The majority of HS are too large and too polar to be analyzed effectively. Thermal and chemical degradation coupled with derivitization can be highly efficient in releasing smaller volatile compounds. Derivitization protects the fragments from reaction and reduces hydrogen bonding (reducing polarity of the molecule). Tetramethylammonium hydroxide thermochemolysis–gas chromatography/mass spectrometry (TMAH-thermochemolysis GC-MS) is a relatively new technique that uses derivitization. Because of the reaction conditions and derivitization step, this technique samples a unique set of components from HS, which are complementary to those found using other degradative techniques. TMAH chemolytically hydrolyzes and methylates esters and ether linkages, releasing fatty acid methyl esters of plant and microbial origin. Thus, TMAH coupled with pyrolysis gas chromatography–mass spectrometry (Py-GC/MS) and ^{13}C-NMR has been used to show the contribution of microbially derived fatty acids to HS.

Chemical Perspective on Microbial Role in Humification

Based on the information obtained from newer chemical methods, it is now widely believed that humic materials are less aromatic in nature than previously thought and are dominated by aliphatic components. The source of the aliphatic components is thought to be plant-derived polymers such as cutin, cutan, or suberin. However, it is now also thought that microbial biomass components such as lipids or lipopolysaccharides may play a part in HS genesis. Aliphatic components can be due to incorporation of microbial materials as well as microbial alteration of primary carbon structures. Thus newer methods indicate a greater role for microbes and microbial activity than has been previously assumed (Table 1).

A top-down approach to genetic inference suggests that microbes are actively involved in HS genesis at several levels. From a strictly chemical (in contrast to an ecologic or biological) viewpoint, there are several microbially mediated 'pathways' by which HS substances may form (Figure 4):

1. Degradation and abiotic condensation: Microbial degradation products and metabolites (amino sugars, semiquinones, or phenolic acids) undergo spontaneous condensation reactions to form complex,

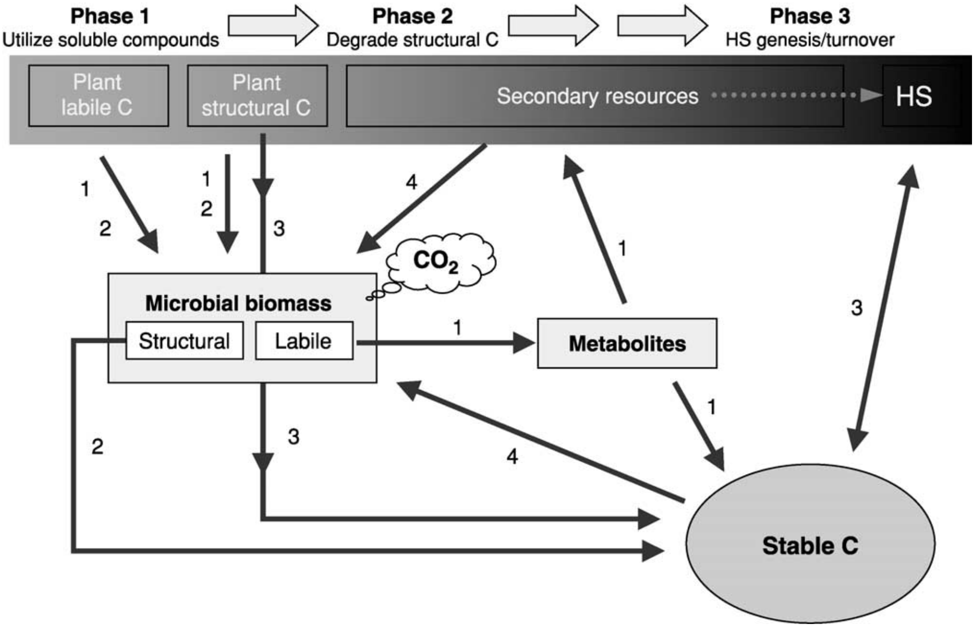


Figure 4 Pathways of microbial involvement in humification. From a strictly genetic standpoint, microorganisms participate in humification via four possible avenues: (1) degradation and abiotic condensation of metabolites and biomass components; (2) degradation of primary resources and resynthesis into refractory components; (3) selective preservation; (4) direct participation in HS production and degradation.

random, macromolecular humic materials. This pathway has long been recognized as one of the primary ways that HS form, particularly in aquatic environments;

2. Degradation of primary resources and resynthesis into recalcitrant components: Microbes consume plant litter, incorporating the carbon into their biomass. Upon death, their cell wall and membrane components are nascent humic substances. This mechanism of HS formation has been suggested fairly recently, with the rise in nondegradative techniques for studying HS. In particular, fungal melanins and bacterial lipid materials may be highly recalcitrant;
3. Selective preservation: Microbes degrade readily accessible plant carbon (e.g., small compounds and side chains), leaving only large, resistant structures such as the 'backbone' from lignin or tannin compounds, alkyl carbon from plant cuticles. The selective preservation of carbon from plant lignins and tannins has also long been acknowledged as a primary microbially mediated process in humification;
4. Direct participation in HS production and degradation: Microbes participate directly in humus genesis and turnover via enzymatic activities. This is a critical yet often overlooked role for microorganisms in humification. Through enzymatic activities, microorganisms can directly polymerize subunits into larger, humic molecules. In addition, microbial activity is a dominant means for HS turnover and loss from soil.

These pathways represent the current chemical view of the role microorganisms play in humification. However, the occurrence of a given microbial pathway and the rate of carbon movement along it will depend on microbial activity and community composition.

Microbiological Perspective

Chemical studies of HS have provided a large array of structural information, and, while structure can indicate the properties and reactivity of HS in soil, it cannot indicate the rate and reasons for HS formation. Ultimately, humification is a consequence of microbial growth and activity. The organisms performing humification are surrounded by carbon at all stages along a decay continuum (Figure 5). Their requirements for energy and survival lead to degradation and processing of soil carbon. The result is humification.

Microorganisms in the soil are influenced by, and interact with, their entire chemical and physical environment rather than with specific chemical compounds in an isolated HS fraction. Thus an understanding of humus genesis and rate of formation requires an understanding of the reasons why humification occurs. This falls within the realm of microbial ecology. From a microbiological standpoint, humification occurs in three phases: (1) rapid initial decomposition of primary plant residues; (2) slow decomposition of plant structural

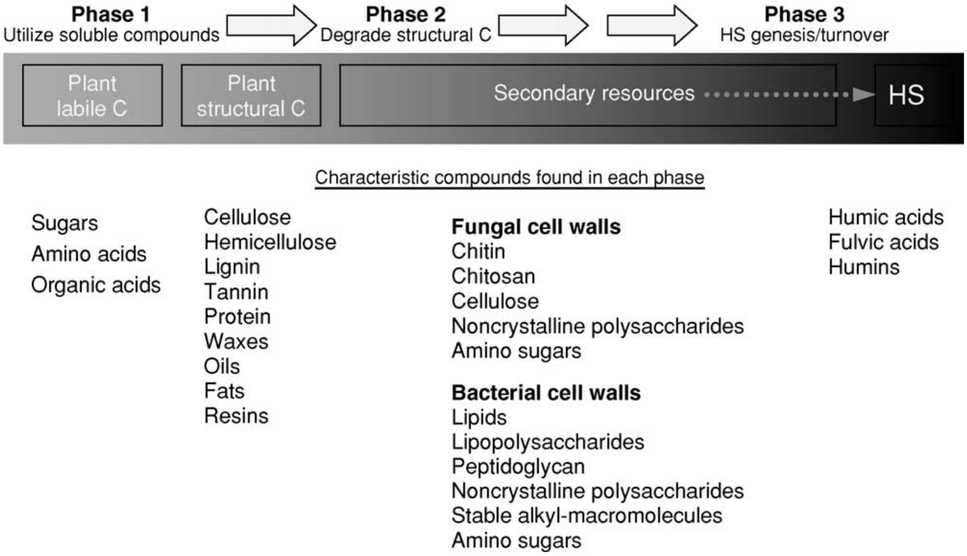


Figure 5 Microorganisms exist surrounded by a continuum of organic matter at all stages of decay and humification. These include primary resources (plant-derived), secondary resources (microbial biomass components) and partially or fully humified organic materials. From a microbial perspective, humification can be thought to occur in three phases that move carbon along the decay continuum: phase 1 is utilization of soluble plant compounds (e.g., amino acids and sugars); phase 2 is degradation of plant structural carbon (e.g., polymeric compounds such as cellulose or hemicellulose); phase 3 is the direct genesis or turnover of humic materials.

components; and (3) alteration of soil organic carbon and HS genesis.

Phase 1: Initial Decomposition of Primary Resources

The first phase in humification is decomposition of labile, plant-derived carbon (litter, or 'primary resources'). Decomposition of terrestrial plant litter is the mechanism by which carbon and nutrients are returned to either the atmosphere or to a plant-available state. Microorganisms utilize soluble, low-molecular-weight compounds ('labile' substrates such as sugars, amino acids, and organic acids) for energy generation and biomass production. Due to their small size, these substrates are readily taken into the cell and metabolized. Because labile substrates (such as glucose) are often low in available nutrients, degrading organisms immobilize nutrients such as nitrogen, phosphorus, or calcium in their biomass during this first phase of decomposition. The microbial community able to utilize these substrates is highly diverse; nearly all organisms will use labile substrates in preference to larger, more complex forms of carbon.

Phase 2: Slow Decomposition, Energy Limited

Degradation rate slows once labile sources of carbon have been utilized, and primarily lignin-encrusted cellulose and hemicelluloses remain. These plant structural compounds are significantly larger in size than water-soluble or labile components from phase 1 and cannot be taken into the cell for degradation. Instead, they must be depolymerized extracellularly prior to uptake and metabolism (Figure 6). Plant structural components surround proteins and other nitrogen-containing resources, making them physically inaccessible. To access nitrogen, organisms must attack the structural components. This requires systems of multiple enzymes, and often multiple organisms (a microbial 'consortium'). In comparison with utilization of labile carbon, there are fewer organisms capable of carrying out depolymerization reactions. Many fungi and eubacteria, such as actinomycetes, are important during this phase of humification. They are capable of producing a variety of cellulolytic and lignolytic enzymes.

Phase 3: Direct Genesis and Degradation

The final phase of humification is direct genesis or degradation of HS. In contrast to phases 1 and 2, microbial contribution to phase 3 is not only through metabolic activity, but also through incorporation of metabolites and biomass components into HS.

The macromolecular random structure of HS makes them highly resistant to microbial enzymatic attack. Microbial enzymes are generally designed for regular repeating carbon units and/or binding sites. Thus, depolymerization and oxidation of HS in soil generally occur incidentally (usually by 'cometabolism'). Cometabolic degradation is oxidation of HS via enzymes produced for another purpose, such as decomposition of plant structural carbon. It yields no

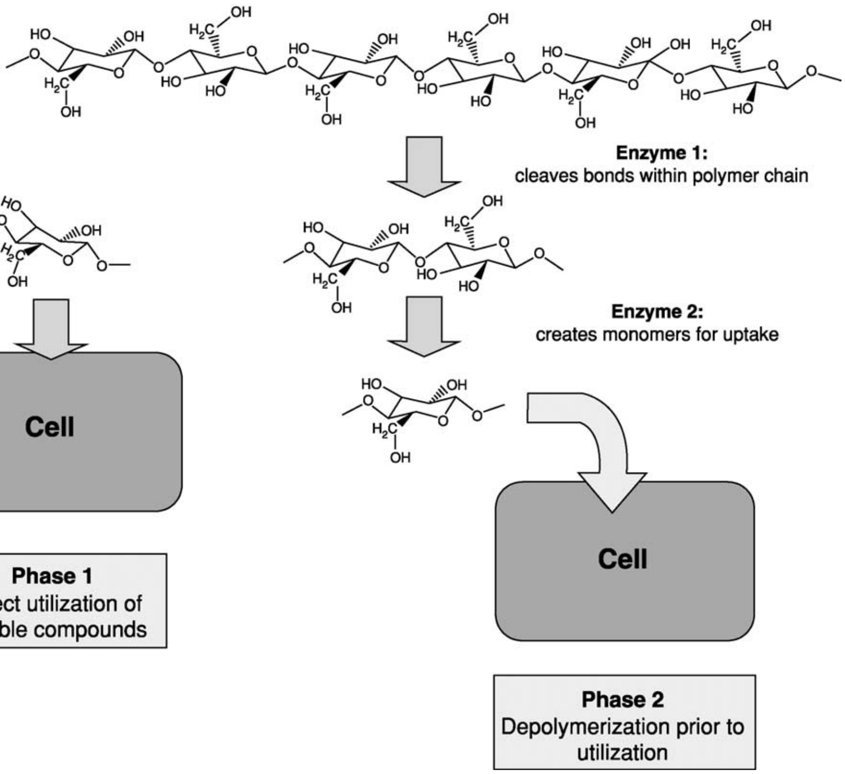


Figure 6 In phase 1 of humification, simple monomeric compounds are taken directly into the cell and metabolized. There are a large array of organisms capable of utilizing labile carbon monomers. In phase 2, plant structural components must be depolymerized prior to utilization. This requires a suite of enzymes and often more than one organism working together in a 'consortium.' Fewer microorganisms are capable of utilizing polymeric compounds.

energetic benefit for the microorganism; rather, an alternate energy source is required (e.g., simple sugars leaked from plant roots). A second source of 'accidental' HS degradation is via enzymes stabilized on mineral surfaces. These can remain active for years, oxidizing and altering HS. Soil mineralogical composition can greatly affect the stability and longevity of this cell-free enzyme pool.

In some circumstances, microorganisms can directly generate and polymerize aromatic compounds to form HS. For example, under low-oxygen conditions, it has been found that microorganisms can make polyaromatic hydrocarbons such as phenanthrene, fluoranthrene, pyrene, and chrysene. It is not known whether this is incidental or serves a biological purpose.

Finally, some of the most important recent studies of humification focus on the incorporation of microbial components into humic materials. First, microbial metabolites and cell components can condense abiotically to form HS. For example, phenolic compounds that are released into the soil from litter or microbial activity are highly reactive and spontaneously undergo nonenzymatic chemical reactions to form more complex molecular structures. In addition, techniques such as TMAH, NMR, and pyrolysis on soil chemical fractions have shown that microbial metabolites and cellular material regularly contribute to HS. For example, the contribution of microbial fatty acids increases with degree of soil organic matter humification and decreasing particle size, suggesting selective preservation of these molecules. In addition, fungal and actinomycete pigments or melanins can be very similar in structure to older HS. Recent ^{15}N-NMR studies also indicate that the majority of organic nitrogen is present in the form of amide and amino N, potentially from preserved microbial proteins. It is well known that microbially derived enzymes can become irreversibly bound to HS in soil. Part of the refractory amide nitrogen may also derive from the bacterial cell wall materials such as peptidoglycan and fungal chitin (Table 2, Figure 7). Lastly, studies of amino sugars have also indicated a critical role for microbial components in HS structures. Muramic acid uniquely originates from soil bacteria, while the compound glucosamine is found in both bacterial and fungal cell walls. Using these marker compounds, it has been possible to show that microbial products accumulate in HS with age and depth in the soil.

Table 2 Some important microbial cell wall components

Cell wall component[a]
N-Acetylglucosamine (amino sugar monomer: glucosamine)
N-Acetylmuramic acid (amino sugar monomer: muramic acid)
Peptidoglycan structure
Chitin
Chitosan

[a]Structures shown in Figure 7.

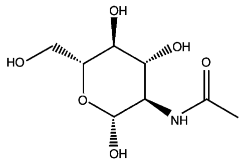
Regulation of Humic Substances Genesis and Turnover

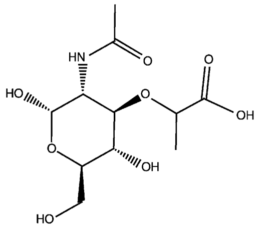
Humification, like all microbially mediated processes, is controlled by factors that affect microbial activity and access to carbon and energy sources. These factors include environmental constraints (temperature, water, pH), nutrient availability, adsorption to surfaces of soil minerals or encryption within a clay mineral coating, and occlusion within micropores in soil structure (Figure 8). In addition, factors such as the resource ‘quality’ available for microbial growth (e.g., plant litter) act as controlling variables.

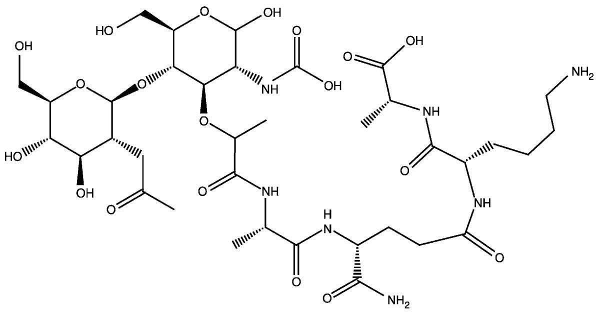
Climate and Soil Type

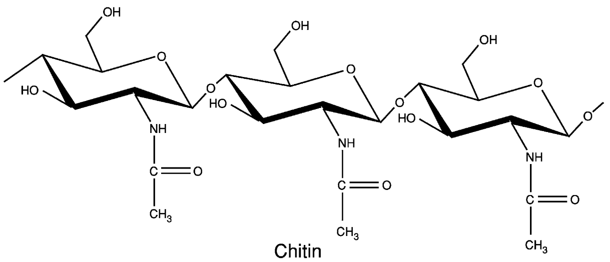
Typical studies of carbon dynamics in soil focus on physical and environmental conditions. Environmental conditions that prevent microbial utilization of C lead to high levels of stored organic matter – as seen in the tundra (where temperature limits utilization) and in peat bogs (where oxygen is limiting). In these situations, microbial enzyme activity limits carbon degradation and humification. For example, monooxygenase enzymes are required for degradation of aromatic structures such as phenolic compounds. These enzymes are inactive in anoxic environments. The carbon in these environments is not humified, but is nevertheless unavailable for microbial utilization. When the environment changes, for example, if the peatlands dry or the climate warms, then the stored carbon may be released rapidly to the atmosphere.

Soil type (e.g., mineralogy and texture) determines physical protection of soil carbon via interaction with soil minerals or within aggregates. Both past and recent work has demonstrated that soils with higher silt and/or clay contents or better aggregate structure tend to have higher levels of carbon. Soils with poorly crystalline minerals such as allophane also maintain high carbon levels (Figure 2). Adsorption to mineral surfaces represents an important process in preventing microbial degradation of stable carbon compounds.


N-Acetylglucosamine


N-Acetylmuramic acid


Peptidoglycan structure


Chitin

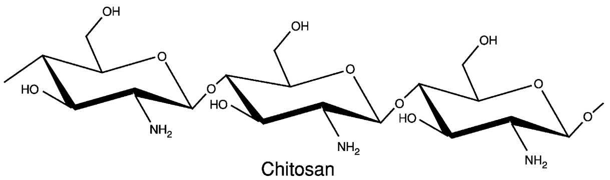

Chitosan

Figure 7 Chemical structures of some important microbial cell wall components. (See Table 2.)

Studies employing bulk dissolved organic carbon (DOC) compounds from soils indicate that the extent of sorptive preservation is related to mineral surface area, but, more importantly, to mineral surface properties. More strongly adsorbed molecules, including intrinsically labile compounds, appear to accumulate in soils. In effect, organic compounds bound to clay

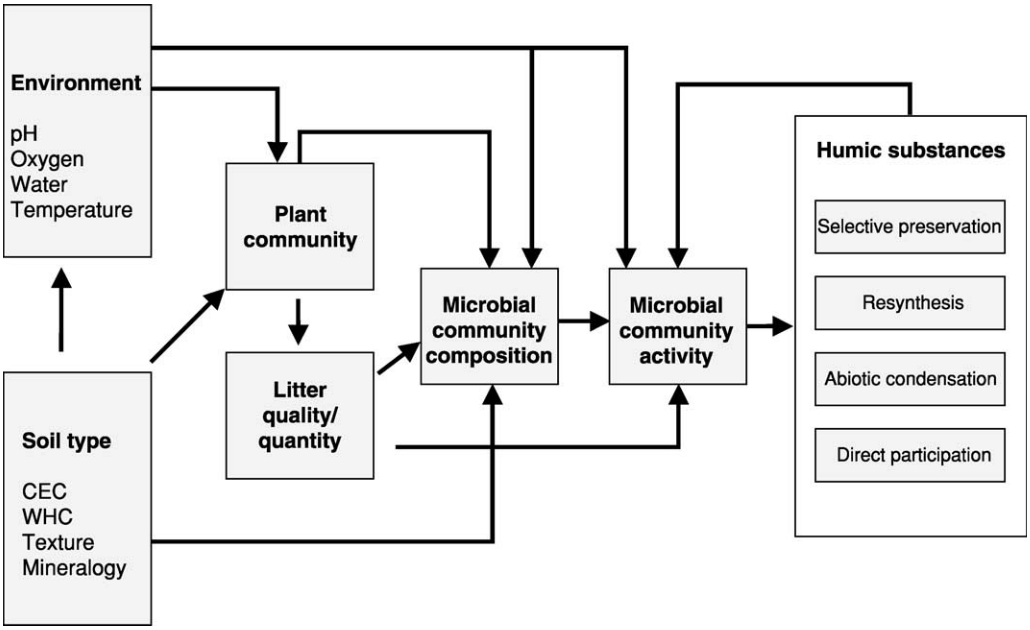


Figure 8 Points of regulation for humification and carbon storage. Physical environment, soil type, and litter quality and quantity all affect microbial community activity and composition, which in turn influence the rate and occurrence of humification. CEC, cation exchange capacity; WHC, water-holding capacity.

minerals are not readily accessible to microbial attack. This slows the rate of humification considerably.

Amount and Chemical Composition of Starting Resources

A second level of control over carbon storage is the importance of litter quality and quantity. Fresh plant litter is degraded at a rate that is proportional to its initial C:N ratio, or perhaps more accurately, proportional to its initial lignin:N ratio. The ratio of C:N or lignin:N is a proxy for substrate quality. Substrate quality controls carbon dynamics via potential energetic gain. Potential energetic gain is determined by oxidation–reduction status, and molecular length and heterogeneity. Highly reduced compounds yield more energy upon oxidation by microorganisms. These types of substrates are utilized preferentially. Molecular length and heterogeneity will also determine whether or not a substrate is preferred. Monomeric compounds (such as glucose) can be taken directly into the cell and require only 'standard' enzymatic machinery for degradation. Polymeric compounds (e.g., cellulose) have the same energetic yield per unit carbon, but require the concerted action of several enzymes. The 'cost' to produce the additional enzymes makes degradation more difficult and therefore less likely. Polymers having irregular or random structures (such as lignin or HS) are also more difficult to degrade than those possessing regular structure. The energetic yield from a random macromolecular compound is low; thus microorganisms will preferentially utilize other carbon sources.

Role of Microbial Community Composition

This topic deserves special mention, because past humification studies have tended to ignore the composition and diversity of the organisms involved. However, the organisms present in the soil can determine the occurrence and rate of humification. A wide range and diversity of organisms participate in phase 1 (simple decomposition). However, phases 2 and 3 require organisms capable of producing somewhat specialized extracellular enzyme complexes. Thus microbial community composition can potentially be a determining factor in the quality and quantity of HS production and turnover in soil (Figure 9).

Microbes, and microbial community composition, impact humification in two ways: (1) via cellular components, and (2) through metabolic activity. Microbes have the ability to take labile carbon and turn it into resistant and/or reactive compounds, but not all microorganisms are equal in this ability. For example, microbial cell wall composition (and recalcitrance) varies widely between Gram-negative bacteria, Gram-positive bacteria, and fungi. If cell walls become nascent humic substances, then the relative abundance of a given wall type should affect the amount of humus produced. Or, since decomposition

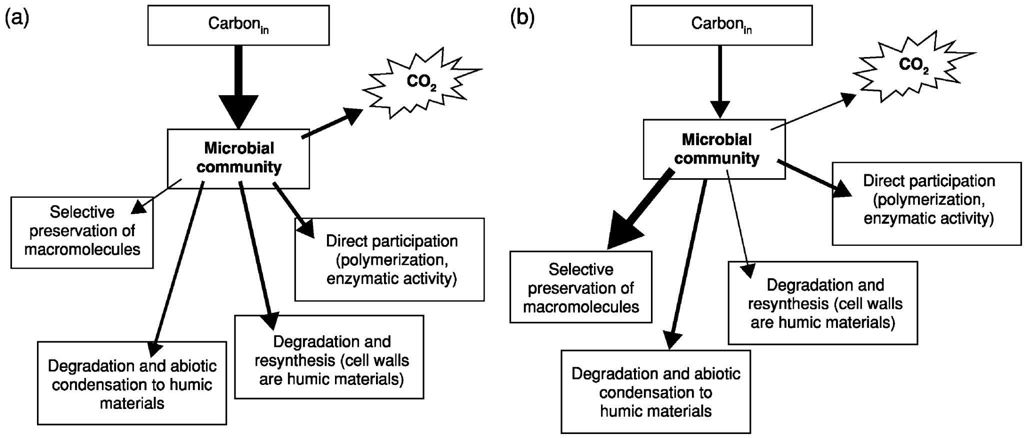


Figure 9 Hypothetical carbon cycle differences for (a) fungally versus (b) bacterially dominated microbial communities. The flow and fate of carbon inputs will vary: fungi degrade incoming plant litter more completely than bacteria, releasing more carbon as CO_2. Because fungal cell walls are highly recalcitrant, humification by degradation and resynthesis may also be high. In contrast, bacteria may contribute more to humification via selective preservation of macromolecular plant litter compounds.

of complex molecules requires a suite of enzymes that not all microbes possess in entirety, the relative importance of selective preservation versus degradation will depend on the genetic makeup of the microbial community. Thus the humification pathway of a given carbon molecule in soil is determined: (1) by the genotype/phenotype of the organism that the C encounters (does it have the ability to use the C?); (2) by the organism's metabolic status (is it starving, or growing, or dormant?); and (3) by the organism's proximity to others (if the initial organism cannot use it, is there one nearby that can?). An excellent example is seen in the forests of the Pacific Northwest USA; the flow and fate of carbon from fallen logs in these wet temperate forests can be entirely determined by microbial community composition. Colonization of logs by wood decay fungi of differing functional types dramatically impacts the resulting carbon cycle. White rot fungi are able to degrade lignin completely to CO_2 and water, and typically leave cellulose untouched. Brown rots are specialists on cellulose, and leave lignin mostly unaltered. White and brown rots do not typically coexist on a single log; the initial colonizer will actively prevent colonization by other fungal species. A log colonized by white rot is reduced to nearly pure cellulose, and one colonized by brown rot is left as a nearly pure source of lignin. Thus at a gross level, the flow and fate of carbon in the system are determined by which fungus colonizes a given log first.

Another example of when microbial community composition might impact humification and carbon storage is in locations within an ecosystem or soil that vary in microbial community composition. For example, forest soils differ from managed and grassland soils; they have highly complex plant litter, and the microbial community is often dominated by filamentous organisms (fungi). Because fungi are generally considered to be more metabolically 'talented' than bacteria, and also have cell walls that are highly resistant to degradation, humification in a forest soil is likely to be a balance between fungal degradation of and contribution to humic materials. The flow of carbon from litter to microbial biomass and into HS will differ from that in a nonfungal-dominated soil (Figure 9). In contrast, the microbial community in a heavily managed soil is likely to be dominated by bacteria, coupled with differing carbon inputs. The process and rate of humification are likely to differ from that in a forest soil.

Finally, even within a given soil, the amount of carbon moving from pool to pool (Figure 2) and the dominant humification pathway (Figure 8) can vary widely. The rhizosphere is generally dominated by Gram-negative bacteria and mycorrhizal fungi, which differ considerably in their ecology (e.g., different metabolic requirements and preferences) from Gram-positive bacteria or saprotrophic fungi. Thus the flow and fate of carbon in the rhizosphere will differ from that in bulk soil (which is dominated by Gram-positive bacteria and saprotrophic fungi).

Summary

Because the generation and turnover of humic substances in the soil is largely a by-product of microbial growth and activity, an understanding of humus genesis must be predicated upon an understanding of microbial community ecology. Factors that control microbial activities and community composition will also control the rate and quality of humification. Ultimately, this understanding of why, where, and how rapidly humification occurs will be critical in our understanding of global carbon sequestration, as well as our general understanding of the nature and properties of the world's soils.

List of Technical Nomenclature

Carbon stabilization	The process of converting carbon readily utilized as a microbial energy source to recalcitrant compounds having a long turnover time. This can be done via chemical alteration (humification) or via environmental or physical restraints on degradation
Degradative approach to study	Macromolecular HS are chemically or physically broken down into various subunits which are subsequently isolated and identified. Structural inferences are drawn based on the identities and yields of degradation products and the reaction processes responsible for their generation
Humic substances (HS)	Conventionally defined as yellow to black-colored, high-molecular-weight, heterogeneous, refractory, naturally occurring organic compounds. They tend to be a range of organic materials that are lumped together almost by default; they cannot be easily classified into discrete categories such as proteins, polysaccharides, or nucleotides. They are often further operationally defined in terms of the methods used to isolate them from soils. Humic acids, fulvic acids, and humin are the three primary fractions, defined on the basis of their solubility in aqueous media, important in soil carbon sequestration due to long turnover times
Humification	Alteration of biotically derived carbon to macromolecular, structurally amorphous forms (e.g., humic and fulvic acids)
Microorganisms or microbes	Organisms not identifiable with the naked eye, including bacteria, fungi, algae, and archaea
Nondegradative approach	Isolated humic substances are analyzed directly, usually without further chemical alteration, utilizing techniques that allow structural inferences to be made
Primary resources	Identifiable plant residues
Rhizosphere	Zone of soil influenced by plant roots. The microbial community in the rhizosphere is generally dominated by Gram-negative bacteria and mycorrhizal fungi
Secondary resources	Residues of soil animals and microorganisms, as well as partially humified compunds
Soil organic matter	Primary and secondary resources, plus dissolved organic carbon, root exudates, and HS. Totality of biologically derived carbon in soil

See also: **Bacteria:** Soil; **Carbon Cycle in Soils:** Dynamics and Management; **Carbon Emissions and Sequestration**; **Fungi**; **Microbial Processes:** Environmental Factors; **Organic Matter:** Principles and Processes; Genesis and Formation; **Organic Residues, Decomposition**; **Pollutants:** Biodegradation

Further Reading

Almendros G and Dorado J (1999) Molecular characteristics related to the biodegradability of humic acid preparations. *European Journal of Soil Science* 50: 227–236.

Chefetz B, Chen Y, Clapp E, and Hatcher PG (2000) Characterization of organic matter in soils by thermochemolysis using tetramethylammonium hydroxide (TMAH). *Soil Science Society of America Journal* 64: 583–589.

Chefetz B, Tarchhitzky J, Desmukh AP, Hatcher PG, and Chen Y (2002) Structural characterization of soil organic matter and humic acids in particle-size fractions of an agricultural soil. *Soil Science Society of America Journal* 66: 129–141.

Gramss G, Ziegenhagen D, and Sorge S (1999) Degradation of soil humic extract by wood- and soil-associated fungi, bacteria and commercial enzymes. *Microbial Ecology* 37: 140–151.

Hatcher P (2001) Modern analytical studies of humic substances. *Soil Science* 166(11): 770–794.

Kaiser K and Guggenberger G (2000) The role of DOM sorption to mineral surfaces in the preservation of organic matter in soils. *Organic Geochemistry* 31: 711–725.

Kogel-Knabner I (2002) The macromolecular organic composition of plant and microbial residues as inputs to soil organic matter. *Soil Biology and Biochemistry* 34: 139–162.

Linhares AA, Linhares LF, and Coelho RRR (1998) Neutral sugars in melanins synthesized by actinomycetes from Brazilian soils. *Biology and Fertility of Soils* 27: 162–167.

Martin JP and Haider K (1971) Microbial activity in relation to soil humus formation. *Soil Science* 111(1): 54–63.

McCarthy P (2001) The principles of humic substances. *Soil Science* 166(11): 738–751.

Oades JM (1989) An introduction to organic matter in mineral soils. In: Dixon JB and Weed SB (eds) *Minerals in Soil Environments*, pp. 89–159. Madison, WI: Soil Science Society of America.

Paul E and Clark F (1996) *Soil Microbiology and Biochemistry.* New York: Academic Press.

Schulten HR and Leinweber P (2000) New insights into organo-mineral particles: composition, properties, and models of molecular structure. *Biology and Fertility of Soils* 30: 399–432.

Stevenson FJ (1994) *Humus Chemistry: Genesis, Composition, Reactions*. New York: John Wiley.

Swift MJ, Heal OW, and Anderson JM (1979) *Decomposition in Terrestrial Ecosystems*. Berkeley/Los Angeles, CA: University of California Press.

Wershaw RL (1993) Model for Humus in Soils and Sediments. *Environmental Science and Technology* 27(5): 814–816.

Zech W, Senesi N, Guggenberger G, Kaiser K, Lehmann J, Miano TM, Miltner A, and Schroth G (1997) Factors controlling humification and mineralization of soil organic matter in the humid tropics. *Geoderma* 79: 117–161.

HYDRAULIC PROPERTIES, TEMPERATURE EFFECTS

S A Grant, US Army Engineer Research and Development Center, Hanover, NH, USA

Introduction

The three soil hydraulic properties upon which temperature has its most pronounced effects are hydraulic conductivity, matric potential, and water-vapor contents. Well-developed theories exist for the effect of temperature on hydraulic conductivity and water-vapor contents. Despite intensive research in the last half of the twentieth century, the mechanism responsible for the effect of temperature on matric potential has yet to be elucidated. Hydraulic conductivity is inversely related to the viscosity of the fluid. The viscosities of fluids, including water, decrease proportionally to the exponent of the reciprocal of temperature – so that hydraulic conductivity increases with increasing temperature. The absolute value of the matric potential decreases linearly with temperature. The mean decrease for the soils studied thus far is $-0.8\%\,\mathrm{K}^{-1}$. In dry soils, water transport in the gas phase can be comparable with that in the liquid phase. The partial pressure of water vapor and its diffusivity increases exponentially with temperature.

Soil-Water Characteristics

Water in unsaturated soils is maintained under negative matric potential. The Young–Laplace equation is generally believed to capture the physics of the phenomenon:

$$\psi = -2\gamma^{\mathrm{lg}}\cos\Theta V_{\mathrm{m,H_2O(l)}}/(rgM_{\mathrm{H_2O}}) \qquad [1]$$

where ψ is matric potential (meters of water); γ^{lg}, surface tension of water (newtons per meter); Θ, the contact angle where the wetting–nonwetting interface meets the soil matrix (radians); $V_{\mathrm{m,H_2O(l)}}$, molar volume of water (cubic meters per mole); g, acceleration due to gravity (meters per second squared); r, the effective pore radius (meters); and $M_{\mathrm{H_2O}}$, molecular weight of water (kilograms per mole). When applied to porous media, the Young–Laplace equation states that at a given matric potential all the pores with effective radii greater than r are empty of the wetting liquid; that is, matric potential increases as the degree of saturation by the wetting liquid decreases. In the hydrology literature, the term cosΘ is used to express the wettability of the porous matrix. The value of cosΘ can vary from 1, signifying completely water-wet, to +1, which signifies completely nonwetting-phase wet.

Empirical relations such as the van Genuchten equation relate soil water content to matric potential:

$$S_{\mathrm{e}} = \left\{\frac{1}{[\alpha\psi]^{\mathrm{n}} + 1}\right\}^{\frac{\mathrm{n}-1}{\mathrm{n}}} \qquad [2]$$

where S_{e} is the water saturation defined by:

$$S_{\mathrm{e}} = \frac{\theta - \theta_{\mathrm{r}}}{\theta_{\mathrm{s}} - \theta_{\mathrm{r}}} \qquad [3]$$

where θ is volumetric water content (cubic meters per cubic meter); θ_{s}, saturated water content (cubic meters per cubic meter); θ_{r}, residual water content (cubic meters per cubic meter); and α (per meter) and n, fitted parameters. Figure 1 presents an exemplary isothermal soil-water characteristic, which is the relationship between matric potential and water content.

Relative Permeability

Once parameterized, the van Genuchten equation can be used to calculate relative permeabilities of the porous medium as a function of degree of saturation:

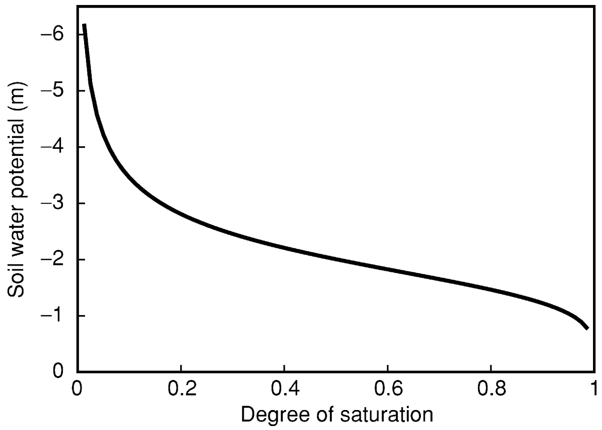


Figure 1 Soil-water characteristic for a loam soil.

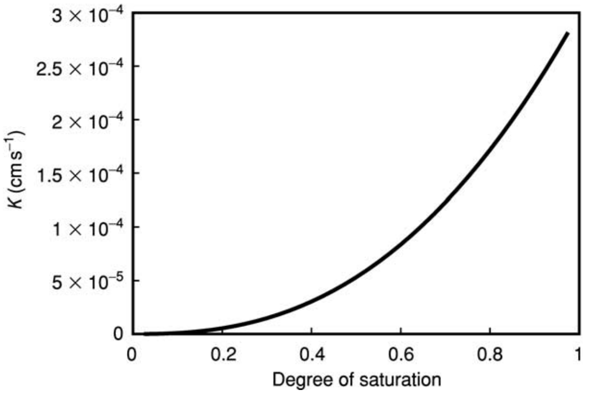


Figure 2 Predicted hydraulic conductivity as a function of degree of saturation predicted by the van Genuchten equation for a loam soil.

$$k_\mathrm{r} = K/K_\mathrm{s} = S_\mathrm{e}^{1/2}\left[\frac{\int_0^{S_\mathrm{e}} \frac{\mathrm{d}x}{\psi(x)}}{\int_0^1 \frac{\mathrm{d}x}{\psi(x)}}\right]^2 \qquad [4]$$

where k_r is the relative hydraulic conductivity (dimension 1); K, unsaturated hydraulic conductivity (meters per second); and K_s, saturated hydraulic conductivity (meters per second). Figure 2 presents the hydraulic conductivity estimated with the van Genuchten equation for the loam soil presented in Figure 1 with an assumed saturated hydraulic conductivity at 25°C of 0.0003 cm s^{-1}.

Effect of Temperature on Soil-Water Characteristics

At room temperature, the matric potential in an 'average' soil decreases 0.8% for every 1 K increase in temperature. The surface tension of water also decreases linearly with temperature. By inspection of eqn [1], soil physicists early in the twentieth century proposed that:

$$\frac{1}{\psi}\frac{\mathrm{d}\psi}{\mathrm{d}T} \overset{?}{=} \frac{1}{\gamma^{\mathrm{ls}}}\frac{\mathrm{d}\gamma^{\mathrm{ls}}}{\mathrm{d}T} \qquad [5]$$

Subsequent experiments have consistently demonstrated that the conjecture presented in eqn [5] is incorrect. The relation underestimates the effect of temperature on matric potential by a factor of approximately 4. Why this is so is not known. Three general mechanisms have been proposed: expansion of trapped gas, solute effects on soil-water surface tension, and temperature-sensitive contact angles. None of these explanations can be reconciled with the available data.

The effect of temperature on matric potential appears to be well described by the relation:

$$\psi(T) = \psi(T_\mathrm{r})\frac{\beta_0 + T}{\beta_0 + T_\mathrm{r}} \qquad [6]$$

where T_r is a reference temperature and β_0 is a constant (K), which for most soils has a value between -350 and -450 K. The available evidence indicates that β_0 is unaffected by soil water content. The matric potential at a reference temperature in eqn [6] can be inserted into the van Genuchten equation to yield:

$$S_\mathrm{e} = \left\{\frac{1}{\left[\alpha\psi\left(\frac{\beta_0 + T_\mathrm{r}}{\beta_0 + T}\right)\right]^{\mathrm{n}} + 1}\right\}^{\frac{\mathrm{n}-1}{\mathrm{n}}} \qquad [7]$$

Figure 3 presents an example of the matric potential–saturation–temperature relationship predicted by eqn [7] for a loam soil. The effect of increasing temperature is to decrease the matric potential gradients within the soil due to differences in water content.

Effect of Temperature on Hydraulic Conductivity

Hydraulic conductivity has been found to be directly proportional to liquid density, the reciprocal of liquid viscosity, and the square of the mean grain diameter. This result leads to the assumption that hydraulic conductivity is separable into distinct contributions owing to the structure of the porous matrix, and the density and viscosity of the fluid. Symbolically, this yields:

$$K = \frac{k\rho g}{\eta} \qquad [8]$$

where ρ and η are the density and viscosity of the liquid (kilograms per cubic meter and pascal-seconds, respectively), g is the gravitational constant (meters per second squared), and k is the intrinsic permeability of the porous matrix (meters squared). The density

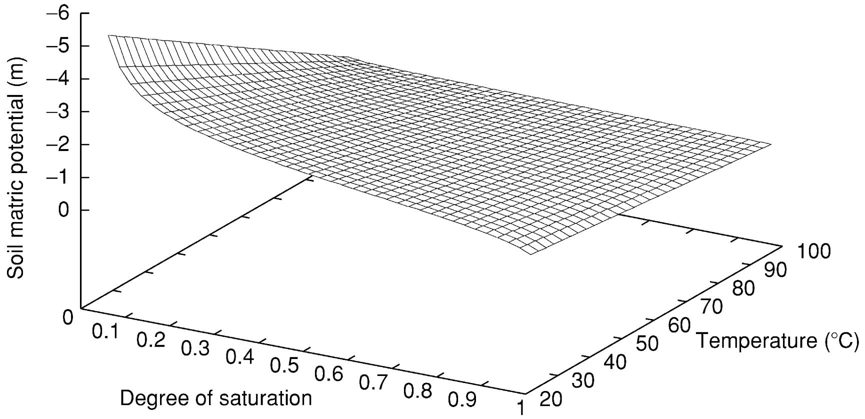


Figure 3 Matric potential as a function of both degree of saturation and temperature for a loam soil.

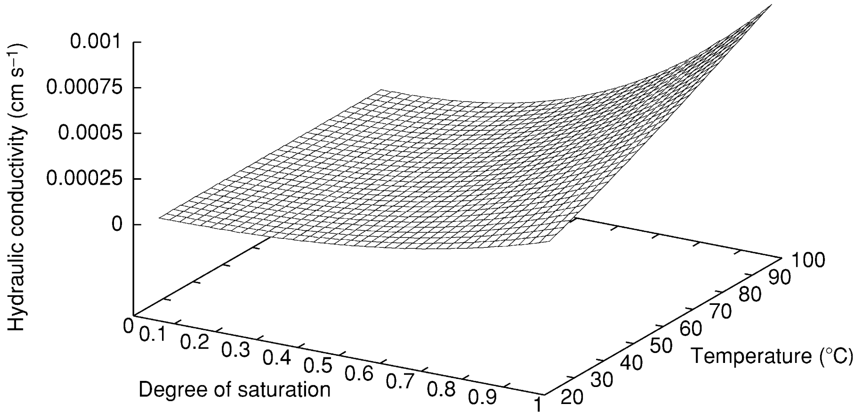


Figure 4 Hydraulic conductivity as a function of both degree of saturation and temperature for a loam soil assuming that the saturated hydraulic conductivity at 25°C is 0.003 cm s^{-1}.

of water changes little between its conventional melting and boiling temperatures, and the approximate value of 1000 kg m^{-3} can be used. The viscosity of water can be fitted to the Vogel–Tamman–Fulcher (VTF) equation:

$$\eta_l = \eta_0 \exp\left(\frac{B}{T + T_0}\right) \qquad [9]$$

yielding the following parameter estimates: $\eta_0 = 31.8903\ (\pm 1.3059)\ \mu\mathrm{Pa\,s}$, $B = 479.9\ (\pm 11.9737)\ \mathrm{K}$, $T_0 = -154.05\ (\pm 1.7891)\ \mathrm{K}$. This equation has a maximum error of −1.1% at 100°C. Figure 4 presents the estimated hydraulic conductivity of the loam soil again assuming that the saturated hydraulic conductivity at 25°C is equal to 0.0003 cm s^{-1}.

Water and Energy Transport in a Nonisothermal Soil

There is a general model of liquid and vapor water and energy transport in soil (*See* **Thermal Properties and Processes**):

$$\frac{\partial \theta}{\partial t} = \nabla \bullet (D_T \nabla T) + \nabla \bullet (D_W \nabla \theta) - \frac{\partial K}{\partial z} \qquad [10]$$

$$C_v \frac{\partial T}{\partial t} = \nabla \bullet (\kappa \nabla T) - L \nabla \bullet (D_W \nabla \theta) \qquad [11]$$

where t is time (seconds); D_T, thermal water diffusivity (meters squared per second per Kelvin), D_W, water content-based water diffusivity (meters squared per second); z, depth (meters); C_v, volumetric heat capacity (Joules per Kelvin per cubic meters); κ, apparent thermal conductivity of the soil (watts per meter per Kelvin); and L, latent enthalpy of vaporization (Joules per cubic meter). The total volumetric water content is the sum of the liquid (θ_l, in cubic meters per cubic meter) and gas water contents (θ_v, cubic meters of liquid-water equivalent per cubic meter of soil).

The thermal water diffusivity can be disaggregated into its liquid- and gas-phase contributions to:

$$D_T = D_{lT} + D_{vT} \qquad [12]$$

where

$$D_{\mathrm{lT}} = -\frac{K_s k_r \rho}{\eta} \frac{\partial \psi}{\partial T} \qquad [13]$$

and

$$D_{\mathrm{vT}} = -(\phi - \theta_{\mathrm{l}}) D \frac{\partial \theta_{\mathrm{v}}}{\partial T} \qquad [14]$$

where D is the gas diffusion coefficient (meters squared per second).

By eqn [5], the temperature derivative of the matric potential is:

$$\frac{\partial \psi}{\partial T} = \frac{\psi}{\beta_0 + T} \qquad [15]$$

The vapor water content of the soil is described by:

$$\theta_{\mathrm{v}} = (\phi - \theta_{\mathrm{l}}) \frac{p^{\mathrm{l+g}} M_{\mathrm{H_2O}}}{RT\rho} \exp[\psi g M_{\mathrm{H_2O}}/(RT)] \qquad [16]$$

where ϕ is porosity (cubic meters per meters cubed), $p^{\mathrm{l+g}}$ is the saturated vapor pressure of water (pascals); R, gas constant (joules per Kelvin per mole). The saturated vapor pressure at a particular temperature can be estimated accurately with the relation:

$$\ln\left(p^{\mathrm{l+g}}/p_{\mathrm{c}}\right) = (T_{\mathrm{c}}/T)\left(a_1\tau + a_2\tau^{1.5} + a_3\tau^3 + a_4\tau^{3.5} + a_5\tau^4 + a_6\tau^{7.5}\right) \qquad [17]$$

where T_{c} and p_{c} are the critical temperature and pressure for water, having values of 647.096 K and 22 064 000 Pa, respectively. The reduced temperature τ is equal to $1 - T/T_{\mathrm{c}}$. The coefficients a_1, a_2, a_3, a_4, a_5, and a_6 have values of −7.8595 17, 1.8440 8259, −11.7866 497, 22.6807 411, −15.9618 719, and 1.8012 2502, respectively. The temperature derivative of vapor water content can be obtained by taking the derivative of eqn [15]:

$$\frac{\partial \theta_{\mathrm{v}}}{\partial T} = \left(\exp[\psi g M_{\mathrm{H_2O}}/(RT)] M_{\mathrm{H_2O}} (\phi - \theta_{\mathrm{l}}) \left\{ RT^2 \left(\partial p^{\mathrm{l+g}}/\partial T\right) - p^{\mathrm{l+g}} [g M_{\mathrm{H_2O}} \psi + T(R - g M_{\mathrm{H_2O}} (\partial \psi/\partial T))] \right\}\right) / (R^2 T^3 \rho) \qquad [18]$$

As in eqn [4], the water-content diffusivity can be disaggregated into its gas- and liquid-phase elements:

$$D_{\mathrm{W}} = D_{\mathrm{lW}} + D_{\mathrm{vW}} \qquad [19]$$

where

$$D_{\mathrm{lW}} = -\frac{K_s k_r \rho}{\eta} \frac{\partial \psi}{\partial \theta} \qquad [20]$$

$$D_{\mathrm{vW}} = -(\phi - \theta_{\mathrm{l}}) D \frac{\partial \theta_{\mathrm{v}}}{\partial \psi} \frac{\partial \psi}{\partial \theta} \qquad [21]$$

To evaluate eqn [18], the constitutive equations describing ψ and k as functions of liquid water content and temperature are needed. In all but the driest soils, a negligible error is introduced into eqns [1], [5], and [7] by assuming that $\nabla\theta = \nabla\theta_{\mathrm{l}}$, and $\partial\psi/\partial\theta = \partial\psi/\partial\theta_{\mathrm{l}}$. Equations [19], [20], and [21] can be evaluated exploiting the van Genuchten equation and noting that the derivative of water-vapor content with respect to matric potential is:

$$\frac{\partial \theta_{\mathrm{v}}}{\partial \psi} = (\phi - \theta_{\mathrm{l}}) \frac{p^{\mathrm{l+g}} M_{\mathrm{H_2O}}^2}{R^2 T^2 \rho} \exp[\psi g M_{\mathrm{H_2O}}/(RT)] \qquad [22]$$

Summary

The degree to which the temperature effect on a given soil hydraulic property is understood depends on the property in question. The theory of the effects of temperature on hydraulic conductivity and gas-phase water concentration are well known and yield precise estimates of these effects. The effect of temperature on matric potential can be described well only for those soils for which the matric potentials have been measured at more than one temperature.

List of Technical Nomenclature

$\boldsymbol{\alpha}$	Fitted parameter in van Genuchten equation ($\mathrm{m^{-1}}$)
$\boldsymbol{\beta_0}$	Constant (K)
$\boldsymbol{\gamma^{\mathrm{lg}}}$	Surface tension of water ($\mathrm{N\,m^{-1}}$)
$\boldsymbol{\eta}$	Viscosity ($\mathrm{Pa\,s^{-1}}$)
$\boldsymbol{\eta_0}$	Fitted parameter ($\mathrm{Pa\,s^{-1}}$)
Θ	Contact angle (rad)
$\boldsymbol{\theta}$	Volumetric water content ($\mathrm{m^3\,m^{-3}}$)
$\boldsymbol{\theta_{\mathrm{l}}}$	Liquid water content ($\mathrm{m^3\,m^{-3}}$)
$\boldsymbol{\theta_{\mathrm{r}}}$	Residual water content ($\mathrm{m^3\,m^{-3}}$)
$\boldsymbol{\theta_{\mathrm{s}}}$	Saturated water content ($\mathrm{m^3\,m^{-3}}$)
$\boldsymbol{\theta_{\mathrm{v}}}$	Gas water content ($\mathrm{m^3}$ of liquid-water equivalent per $\mathrm{m^3}$ of soil)
$\boldsymbol{\kappa}$	Apparent thermal conductivity of the soil ($\mathrm{W\,m^{-1}\,K^{-1}}$)
$\boldsymbol{\rho}$	Density ($\mathrm{kg\,m^{-3}}$)

τ	Reduced temperature
ϕ	Porosity ($m^3\,m^{-3}$)
ψ	Matric potential (m of water)
a_1	Fitted parameter
a_2	Fitted parameter
a_3	Fitted parameter
a_4	Fitted parameter
a_5	Fitted parameter
B	Fitted parameter (K)
C_v	Volumetric heat capacity ($J\,K^{-1}\,m^{-3}$)
D	Gas diffusion coefficient ($m^2\,s^{-1}$)
D_T	Thermal water diffusivity ($m^2\,s^{-1}\,K^{-1}$)
D_W	Water-content-based water diffusivity ($m^2\,s^{-1}$)
g	Acceleration due to gravity ($m\,s^{-2}$)
K	Unsaturated hydraulic conductivity ($m\,s^{-1}$)
K_s	Saturated hydraulic conductivity ($m\,s^{-1}$)
k	Intrinsic permeability (m^2)
k_r	Relative hydraulic conductivity
L	Latent enthalpy of vaporization ($J\,m^{-3}$)
M_{H_2O}	Molecular weight of water ($kg\,mol^{-1}$)
n	Fitted parameter in van Genuchten equation
p_c	Critical pressure (Pa)
p^{l+g}	Saturated vapor pressure of water (Pa)
R	Gas constant ($J\,K^{-1}\,mol^{-1}$)
r	Effective pore radius (m)
S_e	Degree of water saturation ($m^3\,m^{-3}$)
T_c	Critical temperature (K)
T_r	Reference temperature (K)
T_0	Fitted parameter (K)
t	Time (s)
$V_{m,\,H_2O(l)}$	Molar volume of liquid water ($m^3\,mol^{-1}$)
z	Depth (m)

See also: **Thermal Properties and Processes**

Further Reading

Bouyoucos GT (1915) Effect of temperature on the movement of water vapor and capillary moisture in soils. *Journal of Agricultural Research* 5: 141–172.

Constantz J (1991) Comparison of isothermal and isobaric water retention paths in non-swelling porous materials. *Water Resources Research* 27: 3165–3170.

Dullien FAL (1992) *Porous Media, Fluid Transport and Pore Structure*. New York: Academic Press.

Hopmans JW and Dane JH (1986) Temperature dependence of soil hydraulic properties. *Soil Science Society of America Journal* 50: 4–9.

Kakaç S, Kilkis B, Kulacki FA, and Arinç F (eds) (1991) *Convective Heat and Mass Transfer in Porous Media*. Norwell, MA: Kluwer Academic.

Nassar IN and Horton R (1992) Simultaneous transfer of heat, water, and solute in porous media. I. Theoretical development. *Soil Science Society of America Journal* 56: 1350–1356.

Philip JR and de Vries DA (1957) Moisture movement in porous materials under temperature gradients. *Transactions of the American Geophysics Union* 38: 222–232.

Raats PAC, Smiles DE, and Warrick AW (eds) (2002) *Environmental Mechanics, Water, Mass and Energy Transfer in the Biosphere*. Geophysical Monograph Series vol. 129. Washington, DC: American Geophysical Society.

Sageev A, Gobran BD, Brigham WE, and Ramey HJ Jr. (1990) The effect of temperature on absolute permeability to distilled water of unconsolidated sand cores. *Proceedings of the 6th Workshop on Geothermal Reservoir Engineering*, pp. 297–300. SGP-TR-50. Stanford, CA: Stanford University Press.

Wilkinson GE and Klute A (1962) The temperature effect on the equilibrium energy status of water held by porous media. *Soil Science Society of America Proceedings* 26: 326–329.

HYDRIC SOILS

G W Hurt, University of Florida, Gainesville, FL, USA

Introduction

The concept of hydric soils is relatively new to soil science, the term 'hydric soil' having been coined in the late 1970s to designate soils that have formed in wetlands. Only since the 1980s, in response to increased concern over the status of wetlands, have these seasonally wet soils received enough attention to merit a special designation of their own. Hydric soils are soils that support or are capable of supporting wetland ecosystems, i.e., soil modifications are not needed to maintain or restore a wetland. In the case of drained hydric soils, only removal of hydrologic modifications are needed to restore wetlands. Conversely, nonhydic soils do not support, nor are they capable of supporting wetland ecosystems; soil modifications are needed to create a wetland on nonhydric soils.

Hydric soils are most often defined as soils that have formed under conditions of saturation, flooding, or ponding long enough during the growing season to develop anaerobic conditions in the upper part. Nearly all hydric soils exhibit characteristic morphologies that result from repeated periods of near-surface saturation and/or inundation for more than a few days. Saturation or inundation when combined with microbial activity in the soil causes a depletion of oxygen. This anaerobiosis promotes biochemical processes such as the accumulation of organic matter and the reduction, translocation, and/or accumulation of iron and other reducible elements. These processes result in characteristic morphologies that persist in the soil during both wet and dry periods, making them particularly useful for identifying hydric soils. Hydric soil indicators are formed predominantly by the accumulation or loss of iron, manganese, sulfur, or carbon compounds.

Hydrology of Hydric Soils

Hydric soils form where water low in oxygen accumulates in the near surface of a soil or on the surface because water cannot leave the soil as fast as it enters. The difference between water inputs and outputs is storage, where the inputs are precipitation (P), overland flow onto the soil (F), and groundwater discharge (D); outflows are evapotranspiration (ET), runoff from the soil (R), and infiltration into lower soil horizons (I); and storage (S) is the water remaining in the soil or on the site:

$$(\mathrm{P}+\mathrm{F}+\mathrm{D})=(\mathrm{ET}+\mathrm{R}+\mathrm{I})+(\mathrm{S}) \qquad [1]$$

If storage is great enough for a long-enough duration, a hydric soil forms.

Hydric soils form in highly complex wetland ecosystems under a variety of soil properties such as texture, hydrologic conductivity, and bulk density, and a wide variety of hydrologic and other climatic conditions and geomorphic settings; and eqn [1] is difficult if not impossible to quantify for all these conditions. A hydric soil can form in any soil material. It has therefore become convenient to determine the hydric status of a soil based on its morphology. The morphology of hydric soils is most often termed 'hydric soil indicators,' which were developed after observing and describing soil morphology of literally thousands of soils that supported wetland ecosystems. Some of the publications that explain morphologies of hydric soil field indicators are listed in Table 1.

Table 1 Publications of field indicators of hydric soils

Date	*Publication*
1990	Hurt GW, Henderson WG Jr, and Puckett WE. Field identification of hydric soils. In: Carlisle VW (ed.) *Hydric Soils of Florida Handbook*. Florida Association of Professional Soil Classifiers and Florida Cooperative Extension Service. Gainesville, FL: University of Florida Press
1992	Hurt GW and Puckett WE. Proposed hydric soil criteria (indicators) and their field identification. In: Kimble JM (ed.) *Proceedings of the Eight International Soil Correlation Meeting (VII ISCOM): Characteristics, Classification, and Utilization of Wet Soils*. Lincoln, NE: USDA/SCS
1995	Hurt GW, Watts FC, and Henderson WG Jr. Soil and water relationships of Florida's ecological communities and hydric soil indicators. In: Carlisle VW (ed.) *Hydric Soils of Florida Handbook*, 2nd edn. Gainesville, FL: Florida Association of Environmental Soil Scientists
1996	Hurt GW, Whited PM, and Pringle RF (eds) *Field Indicators of Hydric Soils in the United States* (Version 3.2). Fort Worth, TX: USDA–NRCS
1997	Mid-Atlantic Hydric Soil Committee. *Field Indicators of Hydric Soils in the Mid-Atlantic United States*. EPA Region III. Philadelphia, PA
1998	New England Hydric Soil Technical Committee. *Field Indicators for Identifying Hydric Soils in New England*, Version 2.0. New England Interstate Water Pollution Control Commission. Wilmington, MA
2002	Hurt GW, Whited PM, and Pringle RF (eds) *Field Indicators of Hydric Soils in the United States* (Version 5.0), Fort Worth, TX: USDA–NRCS

Important Chemical Reactions

The morphologies used to describe field indicators of hydric soils (Table 1) are created by oxidation–reduction (redox) chemical reactions that occur when a soil is anaerobic and chemically reduced. Most redox reactions produce soil colors and odors indicating that they have occurred. A reduced soil is one in which redox reactions have caused reduced forms of O, N, Mn, Fe, or S to be present in soil solution. 'Reduced' is a general term that implies that some element(s) or compound(s) in addition to O are present in reduced form. Common reduced forms found in hydric soils include: H_2O, N_2, Mn^{2+}, Fe^{2+}, and H_2S; their oxidized forms are O_2, NO_3^-, Mn^{4+}, Fe^{3+}, and SO_4^{2-}, respectively.

To understand how morphological features of reduction form, it is best to simplify oxidation and reduction processes and consider them to be separate reactions even though the two types of reactions occur simultaneously. Redox reactions are the chemical reactions in soils that allow us to assess the wetness of a soil. Although lithology and the amounts and types of vegetation have a major role in determining soil color and amounts of organic matter content at a regional scale, at a site-specific site, these reactions control soil color, organic matter content, the amount of oxygen in soil water, and the amounts and kinds of N, Mn, Fe, and S in soil water. Understanding redox reactions allow us to construe the morphology of a soil and determine the probable wetness (location in the soil).

Oxidation is the production of electrons. When a soil is aerobic (soil pores mostly filled with air), although there are many chemical reactions occurring, the one of interest to us is oxidation. Oxidation occurs when bacteria and other micro- and macroorganisms (collectively called microbes) decompose organic matter. This decomposition of organic matter produces substances that donate electrons and hydrogen ions (protons) which are then available for reduction. Organic matter decomposed by microbes produce electrons (e^-) and protons (H^+). The chemical reaction (oxidation) explained above is the reaction that occurs most of the time in soils. Oxidation is the norm; however, when a soil starts to become wet and eventually saturated (most all soil pores filled with water), reduction reactions begin at microsites and, if the soil is wet enough for a long enough duration, spread through the soil. Reduction is the reverse of oxidation: oxidation produces electrons; reduction consumes electrons. In most soils oxidation and reduction reactions are occurring simultaneously at different microsites.

Reduction of Oxygen

Even after a soil becomes saturated, O is still in the water that is in soil pores. As long as air is present in soil water, all electrons produced by organic matter decomposition are consumed by O to make water:

$$24e^- + 6O_2 + 24H^+ = 12H_2O \qquad [2]$$

This is an aerobic reduction reaction and is the first reducing reaction. When most of the O has been reduced, a soil is said to be anaerobic; As long as there is O in soil water, the reaction shown in eqn [2] continues. Once all the O is gone and can no longer enter the soil because it is waterlogged (saturated), the soil is anaerobic.

After the loss of O (O_2) to form water (H_2O), anaerobic microorganisms that are able to use substances other than O dominate the microbial community. Anaerobiosis results in the reduction of nitrates (NO_3^-) to form N (N_2) followed by Mn from manganic (Mn^{4+}) to manganous (Mn^{2+}), Fe from ferric (Fe^{3+}) to ferrous (Fe^{2+}), sulfates (SO_4^-) to form hydrogen sulfide (H_2S), and carbon dioxide (CO_2^-) to form methane (CH_4).

Reduction of Nitrogen

N reduction may explain the increased interest in wetland retention that occurred in the late 1960s and early 1970s. Nitrates in well water do little damage to most humans. The gastric acid in most stomachs generally kills the bacteria that might otherwise reduce the nitrate to nitrite. For infants less than approximately 2 months old, the nitrate is reduced to nitrite and binds with the infant's hemoglobin. The result is methemoglobinemia, commonly called blue baby syndrome, in which red blood cells carry less oxygen than normal and arrested brain development or death result. Hydric soils in wetlands reduce the nitrate to N_2. This is one of the denitrification reactions:

$$10e^- + 12H^+ + 2NO_3^- = N_2 + 6H_2O \qquad [3]$$

There are similar denitrification reactions that produce nitrite and ammonium (NH_4^+). Excess nitrate in wetlands can lead to eutrophication.

Iron and Manganese Reduction

MnO_2 (usually black in color) and $Fe(OH)_3$ (usually reddish in color) are solids which coat soil particles or other soil surfaces (peds, pores, etc.). Mn^{2+} and Fe^{2+} are colorless and in solution; they are dissolved Mn and Fe ions that move through soil water. Soils that are oxidized most all of the time are usually red in color, and soils that are reduced most all the time are usually grayish in color. Soils with intermittent wetness can be a mixture of reds and grays; these are the soils that are the most interesting, because they occur near hydric soil boundaries:

$$2e^- + 4H^+ + MnO_2 = Mn^{2+} + 2H_2O \qquad [4]$$

$$4e^- + 6H^+ + Fe(OH)_3 = Fe^{2+} + 3H_2O \qquad [5]$$

Sulfur Reduction and Methane Production

When H_2S is produced the soil has a smell similar to rotten eggs. Once all the Mn and Fe has been reduced (Table 2), E_h (redox potential) will continue to fall, and when it reaches −150 mv, SO_4^{2-} anions may be reduced to H_2S gas; and if the E_h continues to drop, CO_2 will be reduced to methane gas (CH_4). These reactions occur only in very wet conditions:

$$8e^- + 10H^+ + SO_4^{2-} = H_2S + 4H_2O \qquad [6]$$

$$8e^- + 8H^+ + CO_2^- = CH_4 + 2H_2O \qquad [7]$$

Redox Potential

Redox potential is a voltage differential as measured between a platinum electrode and a reference electrode both in contact with the soil solution. The voltage is related to the chemical composition of the soil solution (reduced Fe, oxidized Fe, O content, etc.). The voltage reading must be converted to E_h by adding a correction of approximately 200 mV. The correction factor varies with temperature and the type of reference electrode and ranges from approximately 197 to 260). The various reduction reactions (explained above) occur when the soil solution is at various redox potentials. These reactions also vary with pH and mineral species present. One suggested correlation between the reduction reactions and E_h is presented in Table 2.

Figure 1 shows specific E_h/pH lines for O and Fe with the E_h values at pH 7 of +350 and +120, respectively, adjusted with a slope of −60 for each pH value less than pH 7. If E_h values measured at specific pHs plot above the upper line O is present; if below, Oxygen is reduced. If E_h values measured at specific pHs plot above the lower line, Fe is oxidized; if below, Fe is reduced.

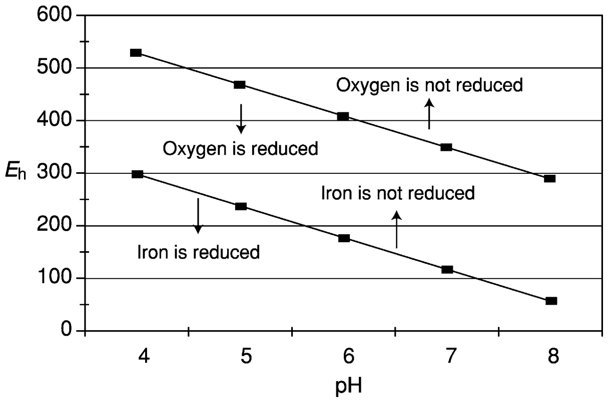


Figure 1 Determination of oxidation–reduction status for oxygen and iron. E_h, redox potential.

Table 2 Reduction reactions and E_h at pH 7

Reduction reaction	E_h *(mV)*
$24e^- + 6O_2 + 24H^+ = 12H_2O$	+350
$10e^- + 12H^+ + 2NO_3^- = N_2 + 6H_2O$	+220
$2e^- + 4H^+ + MnO_2 = Mn^{2+} + 2H_2O$	+200
$e^- + 3H^+ + Fe(OH)_3 = Fe^{2+} + 3H_2O$	+120
$8e^- + 10H^+ + SO_4^{2-} = H_2S + 4H_2O$	−150
$8e^- + 8H^+ + CO_2^- = CH_4 + 2H_2O$	−250

Factors that Control Redox Potential

The factors that control reduction in soils and redox potentials are: (1) organic matter must be present (source of electrons). Good sources of organic matter include dead roots, microbial biomass, plant debris such as leaves, and pieces of roots that come from live roots. Roots even excrete a substance similar to sap upon which bacteria and other organisms thrive. Dissolved organic carbon (material previously excreted by an organism) is utilized by other organisms; (2) air must not enter the soil (soil is saturated); (3) organisms must be decomposing organic matter. This activity is termed 'microbial active.' For most of the world, wet soils are microbial active when air temperature is above approximately 5°C. Reduction, which occurs only when microbes are active, is believed to occur in tundra and Antarctica when the air temperature is below 5°C because of the presence of ethylene glycol (antifreeze) in microbes; (4) dissolved O in the soil water must be removed.

Reduction occurs where water moves through soil slow enough that O is removed and occurs best where water does not move through the soil and most all pores are filled with water. Reduction does not occur where water flows too quickly to allow O to be removed. O must be removed before reduction of other elements can occur. Soil may be saturated but not reduced if: (1) organic matter content is low, (2) microbial activity is lacking, and/or (3) water moves too quickly (O is not removed).

Hydric Soil Indicators

Redox reactions leave signs in the soil that they have occurred as shown in Table 3.

Table 3 Examples of features formed by reducing conditions

Reduced element	*Morphological feature formed*	
	Feature type	*Examples*
Oxygen	Organic (C)-based	Organic soil material Mucky mineral soil material Some black A horizons
Nitrogen	None	None
Manganese	Mn-based	Mn concentrations (mostly black) in a dominantly depleted matrix
Iron	Fe-based	Fe concentrations (mostly red) in a dominantly depleted matrix Reduced matrix
Sulfur	S-based	Rotten egg odor
Carbon	None	None

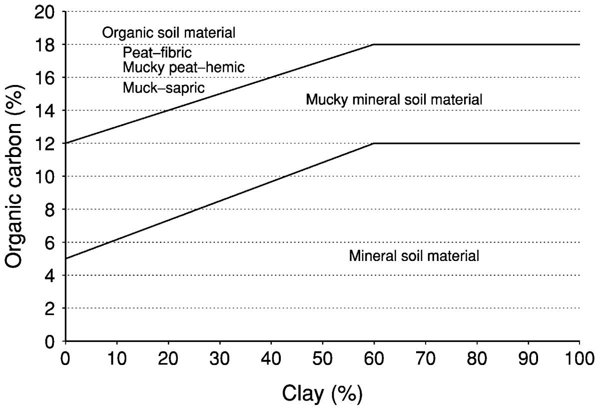


Figure 2 Percentage of organic carbon required for organic soil material, mucky mineral soil material, and mineral soil material as it relates to percentage of clay content.

Carbon Features

Carbon-based features occur when decomposition (of carbon) is slower than production and occur because oxygen is absent. Carbon features are widespread and must be used to identify reduction in soils that contain little or no iron/manganese. Organic-based features consist of one of three kinds of materials: organic soil material and mucky mineral soil material (Figure 2), or mineral soil material with a black color. An exact amount of carbon is usually required (Figure 2). Color and carbon content are the results of relatively high amounts of organic material that accumulate under reduced conditions.

N-based features do not exist. There are no known soil morphological features that are the results of the redox reactions of nitrogen.

Iron and Manganese Features

The reduction, translocation, and oxidation of Fe/Mn oxides form Fe/Mn-based features. The types of features are redox concentrations and redox depletions collectively known as redoximorphic features and are formed by the reduction, movement, and oxidation of Fe and Mn compounds. There are three basic kinds of redoximorphic features: redox concentrations, redox depletions, and the reduced matrix. Both Fe and Mn are used, because the two are very similar in soils. In the absence of additional information, the guidelines in Table 4 can be used to identify the oxidized state of Fe and Mn oxides.

Redox concentrations are zones (usually reddish in color) where Fe/Mn is concentrated. There are three types: (1) nodules and concretions, which are extremely firm (hard when dry), irregular-shaped bodies; (2) masses, which are soft bodies within the soil matrix; and (3) pore linings, which are coatings or impregnations on a natural soil surface (root pore, air pore, ped face, crack face). Pore linings are the most reliable evidence of contemporary wetness, and nodules and concretions are the least reliable evidence of contemporary wetness, because most often they reflect prior wetness. Masses may be the results of reduced iron being reoxidized or evidence that that volume of the soil has never been reduced.

Table 4 Color and composition of iron and manganese oxides

Chroma	*Value*	*Dominant composition*
≤2	≤2	Manganese
3	3	Iron and Manganese
≥4	≥4	Iron

Redox depletions are zones (usually grayish in color) where Fe/Mn oxides have been removed. Usually, when a soil has areas of redox depletions, it also has areas of redox concentrations. If something is removed and translocated, it usually shows up somewhere else. A specialized case where this in not true is the ‘reduced matrix.’ A reduced matrix is gray when freshly exposed and turns redder when exposed to air, because the Fe has been reduced and is thereby colorless but still present in soil water. The Fe oxidizes when exposed to air.

If microbes are active, if Fe/Mn oxides are present in the soil, if the soil is saturated, if the water is stagnant, if there is a source of usable organic matter in the soil, and if O and N have been reduced then the electrons produced by microbes decomposing the organic matter will reduce and translocate the Fe/Mn present and redox features will result. Usually the greater the length of the reduced period the greater will be the percentage of the soil mass redox depletions.

Except for a few exceptions that occur on specific landforms, hydric soils with redox features (features formed by Fe/Mn redox reactions) have more redox depletions than redox concentrations. Fe/Mn features are widespread but occur only in soils with Fe/Mn. Soils that have formed in parent material low in Fe/Mn may have features which appear as depletions. In these soils, carbon-base indicators are used to determine their hydric status.

The reduced matrix is a soil matrix that has a low-chroma color *in situ* because of the presence of Fe^{2+}, but whose color changes in hue or chroma when exposed to air as the Fe^{2+} is oxidized to Fe^{3+}. This phenomenon occurs because, although reduction occurs, the reduced iron lacks an outlet (lateral or vertical) and oxidizes when exposed to air.

Sulfur Features

The only sulfur-based feature is the production of H_2S gas; no particular morphology is associated with H_2S production. The presence of H_2S gas is a very positive indicator of reduction. It is rarely identified in the field because it occurs in soils so wet that few question or look for the presence of hydric soils. Mercaptans are often misidentified as hydrogen sulfide. Mercaptans are s-containing organic compounds with the general formula RSH where R is any radical, especially ethyl mercaptan, C_2H_5SH. Swamp gas (methane or natural gas) features do not exist. There are no known soil morphological features that are the results of the redox reactions that result in the formation and dissolution of methane.

Marl Features

One additional morphology not associated with oxidation–reduction reactions but associated with hydric soils is the production of marl at the soil surface. Marl is a limnic material deposited in water by precipitation of calcium carbonate ($CaCO_3$) by algae. Mineralogical analyses have indicated that almost 100% of the soil minerals are calcite that had precipitated from water saturated with calcium bicarbonate. Marl is formed when the land is flooded for several months each year in the summer (hydroperiod) followed by a number of dry months during the winter. During the annual wet season when these lands are flooded, several hundred species of microalgae (periphyton) grow on the surface water (eqn [8]):

$$Ca^{2+} + 2HCO_3 = CaCO_3 = H_2CO_3 \qquad [8]$$

Periphyton growth produces marl. Flooding creates anaerobic conditions and increases the accumulation of organic matter. During the dry period, this organic matter is oxidized. The relative lengths of the dry and wet periods are approximately the same and therefore marl normally contains little organic matter and is mostly white in color. The rate of calcite mud (marl) production in present coastal areas, as measured by radiocarbon dating, has averaged 1.2 cm per 100 years from 1000 years ago to the present. Generally marl soils form a layer 5–120 cm thick above limestone bedrock.

Summary

Oxidation–reduction reactions are mostly responsible for gray colors in soils. These reactions need organic matter, bacteria, and no O to be present in order to occur. The amount of low-chroma color in soil is usually related to how long it has been reduced, not how long it has been saturated; however, some sandy soils are naturally low in Fe and have low-chroma colors not related to reduction. Redoximorphic features form by Fe/Mn reduction, translocation, and oxidation, and the types of redoximorphic features are concentrations and depletions. Soil morphology that results from redox reactions is used to identify hydric soils in the field.

See also: **Factors of Soil Formation:** Climate; **Wetlands, Naturally Occurring**

Further Reading

Carlisle VW and Hurt GW (eds) (2000) *Hydric Soils of Florida Handbook*, 3rd edn. Gainesville, FL: Florida Association of Environmental Soil Scientists.

Environmental Laboratory (1987) *Corps of Engineers Wetland Delineation Manual*. Technical Report Y-87-1. Vicksburg, MS: US Army Engineers Waterways Experiment Station.

Hurt GW and Richardson JL (eds) (1998) *Hydric Soils of the United States: A Guide to Their Recognition*, (CD-ROM). Fort Worth, TX: USDA–NRCS.

Hurt GW, Whited PM, and Pringle RF (eds) (2002) *Field Indicators of Hydric Soils in the United States, Version 5.0*. Fort Worth, TX: USDA–NRCS.

Mausbach MJ and Richardson JL (1994) Biogeochemical Processes in Hydric Soils. *Current Topics in Wetland Biogeochemistry* 1: 68–127.

Munsell® Soil Color Charts (2000). New Windsor, NY: Gretag Macbeth.

Richardson JL and Vepraskas MJ (eds) (2001) *Wetland Soils: Genesis, Hydrology, Landscapes, and Classification*. Boca Raton, FL: Lewis Publishers.

Schoeneberger PJ, Wysocki DA, Benham EC, and Broderson WD (1998) *Field Book for Describing and Sampling Soils*. National Soil Survey Center. Lincoln, NE: USDA–NRCS.

Soil Survey Staff (1999) *Soil Taxonomy: A Basic System of Soil Classification for Making and Interpreting Soil Surveys*. USDA Agricultural Handbook 436. Washington, DC: US Government Printing Office.

Vepraskas MJ (1994) *Redoximorphic Features for Identifying Aquic Conditions*. Technical Bulletin 301, North Carolina Agricultural Research Service. Raleigh, NC: North Carolina State University Press.

HYDROCARBONS

P Kostecki, University of Massachusetts, Amherst, MA, USA
R Morrison, DPRA, Inc., San Marcos, CA, USA
J Dragun, Dragun Corporation, Farmington Hills, MI, USA

Introduction

This article presents a scientific overview of the basic principles involved with remediating sites containing soils contaminated with petroleum hydrocarbons. The first section discusses the physical and chemical composition and properties of petroleum hydrocarbons. The subsequent sections discuss factors affecting the transport and transformations of petroleum hydrocarbons in soil systems, as well as the primary engineering technologies utilized to remediate soils contaminated with petroleum hydrocarbons.

Chemistry of Petroleum Hydrocarbons

Petroleum hydrocarbons and their refined products originate from a highly complex mixture, although they are comprised mainly of only two elements, hydrogen and carbon, which can be arranged in an almost infinite number of structures. There are some distinct rules that govern the arrangement of these atoms. The carbon atoms can be attached or bonded to up to four different atoms, carbon or hydrogen, while the hydrogen atoms are bonded to only one carbon atom. The carbon atoms are bonded to each other and form a structure called the carbon backbone. Using the same number of atoms, many different arrangements of the atoms can be made. Each arrangement constitutes a unique organic compound.

Although each compound has its own unique chemical properties, the vast number of compounds present in crude oil makes it necessary to group the compounds whenever possible. Rather than knowing the chemical properties of all compounds, the general properties of a group or a class of compounds are used to estimate or approximate the expected property for an individual compound. Where more detailed information is needed, the actual properties for a specific compound can be determined.

One method employed to group a hydrocarbon is by the number of carbon atoms that are present in the compound. For example, all of the hydrocarbons that have four carbon atoms can be designated as C_4. The C_4 compounds commonly found in petroleum products include butane, isobutene, 1-butene, 2-butene, and isobutylene. In this example, C_4 can refer to as many as five different compounds, while C_{10} can refer to more than 470 individual compounds.

Compounds from C_1 to C_4 are gases at ordinary temperature. They are used as fuels and in the manufacturing of specialty chemicals, including plastics or polymeric materials. Compounds from C_5 to C_{12} are liquids that will easily turn into gases when heated. These compounds are used as fuels and as cleaning solvents. Compounds from C_{13} to C_{17} are generally liquids and are used as fuels and lubricants. Those compounds greater than C_{17} are usually solids and are used as lubricants, heavy fuels, and as coatings such as tars and asphalt. The solid hydrocarbons usually have the consistency of wax. These solids will often pour or flow when heated or dissolved in a solvent.

A second major grouping is based on the chemical reactivity of the hydrocarbon compounds. The reactivity of these compounds is based on the molecular structure or the actual arrangement of the hydrogen and carbon atoms. The nature of the carbon backbone and the number of hydrogen atoms that are attached to it determine the chemical properties of these compounds. There are three major classifications of hydrocarbons that are found in petroleum products: alkanes, alkenes, and aromatics.

Alkanes

Alkanes are the major constituents of crude oil and the major constituents of most petroleum products. Alkanes are also called saturates or saturated hydrocarbons. The carbon atoms that make up the carbon backbone are linked together to form a chain (linear or branched alkanes), a circle (cyclic alkenes), or a mixture of both. The alkanes are distinguished from the other hydrocarbons by the fact that they are completely saturated with hydrogen. This means that no additional hydrogen atoms can be added to these hydrocarbons without destroying the carbon backbone.

Linear and branched alkanes are sometimes called paraffins. These compounds have an empirical formula of C_nH_{2n+2}. A linear or branched alkane that has four carbon atoms will also have 10 and only 10 hydrogen atoms. Linear alkanes, i.e., alkanes whose carbon atoms are arranged in a straight chain or row, are called normal paraffins, normal alkanes, or *n*-alkanes. The branched alkanes are often called isoparaffins or isoalkanes. The carbon atoms of the carbon backbone of the branched alkanes might look like a T, Y, or X.

The simplest alkane is methane. Methane contains only one carbon atom and four hydrogen atoms. The other alkanes can be made by substituting carbon atoms for hydrogens.

The cycloalkanes are similar to the linear and branched alkanes, except that the carbon backbone has the carbon atoms arranged in a circle. These compounds are sometimes called cycloparaffins and have an empirical formula of C_nH_{2n}. These compounds are also saturated. The carbon atoms that make up the backbone are arranged in circles containing usually five, six, or seven carbon atoms.

Hydrocarbons also exist that are a combination of linear alkanes and cycloalkanes. These compounds have the empirical formula C_nH_{2n}. The fact that they have a cyclic groupin dominates their chemistry and they are usually called cycloalkanes or alkyl cycloalkanes. If these hydrocarbons contain a five- or a six-membered ring, they are called napthenes (not to be confused with naphthalenes.)

Alkenes

The second major group of petroleum hydrocarbon is the alkenes, also known as olefins. These compounds are usually not found in crude oil but are the by-product of a refining process. The alkenes are distinguished from the alkanes by the fact that they contain two less hydrogen atoms than the alkane with the same carbon backbone. Because the alkenes have two fewer hydrogen atoms, they contain a carbon–carbon double bond. This carbon–carbon double bond imparts special chemistry to these compounds different from that shown by the alkanes. The chemistry of the alkenes is usually of little importance to petroleum products since most are simply used as a fuel and burned. The alkenes, especially ethylene and isobutylene, are very important starting materials in the manufacture of specialty chemicals.

The alkenes are considered to be unsaturated because hydrogen can be added (to form the corresponding alkane) without breaking the carbon backbone. Addition of two hydrogen atoms to isobutylene yields isobutene. Alkenes have the empirical formula C_nH_{2n}.

Aromatics

The third basic type of hydrocarbon found in petroleum products is the aromatics. Aromatic hydrocarbons typically contain a six-membered carbon ring and they have an empirical formula of C_nH_{2n-6}. The six-membered ring has three carbon–carbon double bonds. This arrangement of three double bonds within a six-membered ring imparts very special chemistry to these compounds. They can behave in ways that are very different from the other hydrocarbons.

Benzene is the simplest aromatic hydrocarbon. Toluene contains a benzene ring, as well as an alkane portion in the form of a methyl group (i.e., CH_3.) Toluene can also be called a C_1-benzene. Ethyl benzene contains a benzene ring with an attached ethyl (CH_2CH_3) group. There are three different xylenes; each contains two methyl groups attached to a benzene ring. The xylenes are distinguished from each other by the relative location of the two methyl groups on the benzene ring. Ethyl benzene and the three xylenes can be referred to as C_2-benzenes. The higher aromatics such as naphthalene and benzo (*a*) pyrene contain more than one benzene ring. The benzene rings are fused where each ring shares two carbon atoms with its neighboring ring. These latter two compounds are part of a class of hydrocarbons called polynuclear aromatic hydrocarbons (PAH or PNA).

There are two very useful terms that can be used to describe similar yet different hydrocarbons. These terms make it easier to express the complexity of the hydrocarbons that are found in petroleum and in petroleum products: 'homologous series' and 'isomers.'

Homologous Series

A homologous series is a series of compounds that differ from each other by the removal or addition of a common unit or group of atoms. An example of a homologous series of *n*-alkyl benzenes would begin with benzene (C_6H_6) and would include toluene ($C_6H_5CH_3$), ethyl benzene ($C_6H_5CH_2CH_3$), *n*-propyl benzene ($C_6H_5CH_2CH_2CH_3$), *n*-butyl benzene ($C_6H_5CH_2CH_2CH_2CH_3$), and so on. Except for benzene, these compounds all contain a normal alkane attached to a benzene ring. Each succeeding compound contains a normal alkane containing one additional CH_2 unit.

A second homologous series based on benzene could be benzene, C_1-benzene (toluene), C_2-benzenes (ethyl benzenes, *o*-xylene, *m*-xylene, *p*-xylene), C_3-benzenes, and so on. This group differs from the previous in that it includes isoparaffins, as well as more than one alkyl group attached to a benzene ring. Similar series can be made for cyclohexane and naphthalene.

An important homologous series for petroleum products are the *n*-alkanes. The series includes: *n*-C_1 (methane), *n*-C_2 (ethane), *n*-C_3 (propane), *n*-C_4 (butane), *n*-C_5 (pentane), *n*-C_6 (hexane), *n*-C_7 (heptane), *n*-C_8 (octane), *n*-C_9 (nonane), *n*-C_{10} (decane), *n*-C_{11}, *n*-C_{12}, and so on. This latter homologous series is particularly important in the characterization of middle-range petroleum distillates and residual petroleum products such as diesel, fuel oil, jet fuel, kerosene, and bunker oil.

Isomers

Isomers are compounds that contain exactly the same number of atoms, i.e., they have exactly the same empirical formula, but differ from each other by the way in which the atoms are arranged. Examples of isomers with the formula C_8H_{10} are ethyl benzene, *m*-xylene, *p*-xylene, and *o*-xylene. This particular group of isomers is C_2-benzenes, since they all have two carbon atoms attached to a benzene ring.

There can be several isomers for each empirical formula. The total number of possible isomers containing 40 carbon atoms and 82 hydrogen atoms (C_{40} alkanes) is 62 491 178 805 132. Crude oil and its products probably do not contain all possible isomers; however, some highly refined products may contain a significant fraction of the total number of isomers that are theoretically possible. These numbers will provide some insight into the complexity and wide variety of compounds that can be present in petroleum.

Transport of Petroleum Hydrocarbons in Soil

The fundamental principles governing advective (mass) transport of water in soil generally apply to those for hydrocarbon transport (i.e., gravitational and capillarity). Hydrocarbons move through the soil under a three-phase flow condition, displacing air and water. Variations in soil permeability result in a deviation from the gravitationally dominated vertical flow; as the hydrocarbon encounters layers of slightly less permeable materials, or if materials with smaller pores are encountered, it will tend to flow mostly in the horizontal direction until it encounters a path of less resistance.

This conceptual model is more complex, however, because other transport and transformation processes occur. For example, simultaneous processes may occur upon the release of a hydrocarbon at the ground surface:

1. Volatilization of the lighter components, such as benzene, into the atmosphere;
2. Runoff over the soil surface, driven by gravity and affected by the surface configuration and possibly by rainfall, with consequent contamination of nearby surface waters or concentration of the materials in surface depressions;
3. Adherence of the heavier components to the soil surface, possibly resulting in clogging and hydrophobization of the soil;
4. Soil infiltration;
5. Downward and lateral flow within the unsaturated zone in the soil profile;
6. Retention of the hydrocarbon at soil textural interfaces, such as clay, within the soil profile;
7. Retention in soil pores and attachment to soil grains as well as soil organic matter;
8. Volatilization and vapor diffusion within the soil and out of the surface, or possible movement of a vapor cloud vertically into the saturated zone or soil pore water, and possible redissolution into the groundwater at the water table;
9. Chromatographic separation of components within the profile, resulting in selective migration of lighter and less-viscous components;
10. Partial dissolution of soluble or emulsified components within the water phase of the soil;
11. Degradation resulting from both nonbiological (i.e., chemical) and biological processes;
12. Internal drainage or leaching from the soil either within or alongside the water phase toward the water table;
13. Mounding over the water table and divergence of the petroleum hydrocarbon over it, possibly including convergence in cones of depression or drawdown regions at wells or streams;
14. Penetration into the aquifer of the soluble and high specific density compounds;
15. Dispersion and additional migration within the aquifer and eventual appearance in the water supplies.

An example illustrating these possible interactions is useful. If a large volume of gasoline or diesel is released at or near the surface, it will initially tend to infiltrate vertically through the soil. If the volume of release is sufficient to exceed the residual soil-retention capacity, the migration of the hydrocarbon will continue until the fluid reaches the capillary fringe, where it accumulates. As the hydrocarbon is transported through the soil, it may encounter a less-permeable layer, which may cause lateral spreading until a more permeable zone is encountered for the hydrocarbon to move vertically.

On a smaller scale, the distribution of the hydrocarbons in the soil is dependent on the pore size between the soil grains and the pore pressures of the air, water, and hydrocarbons occupying these pore spaces. The permeability of a soil for a hydrocarbon is different from the permeability for water, owing to differences in viscosity. Also, gasoline components will preferentially dissolve from the hydrocarbon. Furthermore, these components may migrate at different velocities through the soil; this phenomenon is called chromatographic separation. Chromatographic separation is often observed in the separation of more mobile compounds such as methyl *tert*-butyl ether (MTBE) moving ahead of the center of mass of a hydrocarbon in groundwater.

Residual Saturation

'Residual saturation' is defined as the fraction of the total soil space that is filled with a liquid due to capillary forces. Residual saturation plays an important role in hydrocarbon transport through the subsurface. As hydrocarbons migrate through soil, a small amount of the total hydrocarbon mass remains attached to soil particles via sorptive forces. The hydrocarbon that is retained by the soil is also known as the immobile saturation. It is dependent on soil-moisture content, soil porosity, and soil texture. Ranges for gasoline and diesel residual saturations for different soils are listed in Table 1.

The residual saturation of a hydrocarbon is also affected by its viscosity. As the viscosity of the hydrocarbon decreases, the residual saturation concentration also decreases. The residual saturation for petroleum products, expressed as concentration (grams per kilogram), in soils is shown in Table 2.

Residual saturation is significant because it remains as a source of contamination for water infiltrating through the soil column that contacts the residual hydrocarbons. Infiltrating water can 'flush' entrapped residual hydrocarbons from the larger pores in a soil.

Residual-saturation contamination by hydrocarbons can occur downgradient from a spill that may be on an adjacent property. This situation occurs when hydrocarbon floating above a water table comes into contact with otherwise clean soils above it due to a rising water table. Subsequently when the water table drops, the previously clean soil retains a certain percentage of residual saturation and thus becomes a source for further contamination. This phenomenon is important to consider when examining soil gas survey data as an indicator of the source of contamination, especially with shallow groundwater. Soils with residual hydrocarbon downgradient of the original spill may volatilize, resulting in the appearance of a release and/or masking of the original source of the contamination.

Table 1 Ranges for gasoline and diesel fuel residual saturation for different soils

Soil texture	*Gasoline (%)*	*Diesel fuel (%)*
Coarse gravel	0.5–2.0	1–4
Coarse sand	1.0–3.0	3–8
Medium sand	2.0–4.0	4–10
Fine sand	3.0–5.0	6–20
Silt	5.0–9.0	10–30

Table 2 Residual saturation for petroleum products in soils (grams/kilogram soil)

Soil type	*Gasoline*	*No. 2 fuel oil*	*Lube oil*	*No. 6 fuel oil*
Coarse gravel	–	0.8	1.6	–
Gravel to coarse sand	–	1.6	3.2	–
Coarse to medium sand	–	2.8	5.6	–
Medium to fine sand	2.0	4.8	9.6	60.0
Fine sand to silt	–	8.0	16.0	–

Estimation of the Distribution of Hydrocarbons in Soil

The soil volume required to immobilize hydrocarbons depends on soil porosity, the physical properties of the hydrocarbons, and the volume of released hydrocarbon.

The American Petroleum Institute (API) has approximated the soil volume required to immobilize a volume of hydrocarbons as residual saturation by:

$$V_s = (\alpha V_C/P)(\mathrm{RS})$$

where V_C is the volume of discharged hydrocarbons, V_s is volume of soil required to attain residual saturation, P is soil porosity, RS is residual saturation capacity, and α is coefficient, depending on the units used (whether metric or other units).

The maximum depth of hydrocarbon penetration in the subsurface by hydrocarbons can be estimated:

$$D = KV_{\mathrm{HC}}/A$$

where V_{HC} is volume of discharged hydrocarbons, A is area of infiltration, and K is a constant that is dependent on the soil's sorptive capacity and hydrocarbon viscosity. Table 3 lists representative K values for three hydrocarbons in soils of various textures.

An approximation to estimate the area over which a spill has occurred is described by:

Table 3 Typical values of K for various soil textures

Soil texture	*K (gasoline)*	*K (kerosene)*	*K (light fuel oil)*
Stone and coarse gravel	400	200	100
Gravel and coarse sand	250	125	62
Coarse and medium sand	130	66	33
Medium and fine sand	80	40	20
Fine sand and silt	50	25	12

$$A_s = 53.5(V_C)^{0.89}$$

where A_s is spill area in square meters, and V_C is volume of discharged hydrocarbon in barrels (1 barrel ≈ 160 l). The actual area may vary from the calculated area, due to climatic, topographic, soil, and hydrocarbon properties.

For a steady-state release of a hydrocarbon (a leaking underground storage tank, piping, etc.) into a soil overlying a relatively impermeable layer under isothermal conditions, the radius (R) of the mass of the bulk hydrocarbon is estimated as:

$$R = (Q/3.142h)^{1/2}t^{1/2}$$

where Q is the rate of hydrocarbon addition to the soil, h is height of the oil film, and t is time after the initial hydrocarbon addition to the soil.

Hydrocarbon Interactions at the Capillary Fringe

Vertical migration ceases when the hydrocarbon encounters water-saturated pores (i.e., the capillary fringe or a perched layer). As the volume of floating product increases, saturation at the capillary fringe becomes sufficient for lateral spreading to occur. A mound above the water table occurs along with a corresponding depression in the water table. Spreading takes place along the groundwater surface as a result of the natural groundwater gradient and the gradient caused by the hydrocarbon accumulating above the groundwater surface. The hydrocarbons will eventually move laterally over the water table in a layer roughly as thick as the capillary fringe.

Some of the hydrocarbon residing in the capillary fringe will slowly volatilize into the vadose zone. The rates of volatilization are controlled by the flow of air. Hydrocarbons resting on the water table tend to compress the capillary zone and lower the water table. The compression depends on the hydrocarbon density. As the specific gravity of gasoline and light hydrocarbons are approximately 0.7–0.8, approximately 70%–80% of the hydrocarbons will lie below the depth of the original water table.

Floating free-phase hydrocarbons that are less dense than water (low-density nonaqueous phase liquids, LNAPLs) may accumulate as a mound and then travel radially in all directions outward from the mound center. This hydrocarbon lens will spread laterally until equilibrium is achieved between the lens thickness and the capillary pressure. As expected, this balance of forces results in a thick layer in fine-grained sediments and a thinner layer in coarse sediments. The center of hydrocarbon mass, however, will move downgradient with respect to groundwater. Volatile components within the LNAPL may also evaporate into the air and diffuse upward, while the higher molecular weight compounds remain in the LNAPL pool. Compounds with low oil–water partition coefficients will tend to diffuse outward and transfer into the aqueous phase, leaving the heavier-weight chemicals within the mound.

Upon entering the capillary fringe, sufficient hydraulic pressure may exist to cause petroleum hydrocarbons to move hydraulically up gradient. This upgradient distance is minimal when compared with the downgradient distance. As the hydrocarbon lens moves laterally, water in the capillary fringe impedes its movement. In the upper section of the capillary fringe, where small amounts of water are present, the hydrocarbons tend to migrate laterally. In the lower section of the capillary fringe, where relatively large amounts of water are present, the bulk hydrocarbons also migrate laterally but more slowly than in the upper portions of the capillary fringe.

The maximum spread of the hydrocarbon lens over the groundwater table can be approximated with the following equation:

$$S = (1000/F)[V - (Ad/K)]$$

where S is maximum spread of the bulk hydrocarbons in meters squared, F is thickness of the hydrocarbons in millimeters, V is volume of bulk hydrocarbon infiltration in meters cubed, A is area of infiltration (meters squared), d is depth to groundwater in meters, and K is a constant dependent on the soil's sorptive capacity for a given hydrocarbon and the viscosity of the hydrocarbon. The vertical thickness of this lens shadows fluctuations of the groundwater table, with hydrocarbon sorption and flushing occurring.

Solubilization of Hydrocarbons into Groundwater

The solubilization of products from a petroleum product results in a chemical composition for the dissolved component in the groundwater that is different than the chemical composition of the initial product. For nonoxygenated gasoline, the most soluble components are benzene, toluene, ethyl benzene, and xylene (BTEX). While these compounds comprise 23–55% of the bulk product in a gasoline, they will comprise from 42 to 74% of the hydrocarbons in the groundwater. Benzene concentrations, for example, are enriched up to 10 times in the groundwater relative to the gasoline. Solubilization experiments have shown that while an unleaded gasoline contains 49% of the aromatic compounds, water in contact with the gasoline contains 95% of these compounds after 30 min of mixing.

For crude oils, the solubility is partially dependent on the API gravity of the crude oil. API gravity is described by the following expression:

$$\text{API} = (141.5 \times \text{density of water at } 60°\text{F} / \text{density of petroleum at } 60°\text{F}) - 131.5$$

Crude oil with a high API is termed 'light' (less dense than water), while a low-API crude is 'heavy.' The solubility of the crude oil in contact with groundwater increases with higher API gravities, because high-API crude oils tend to be rich in low molecular weight compounds with high solubilities. The solubility of crude oil with various API gravities as measured in parts per million (milligrams per liter) total dissolved hydrocarbons is given in Table 4.

The mass of soluble hydrocarbons that actually dissolve into the groundwater is poorly defined. One expression for approximating the amount of dissolved substance originating from free-phase hydrocarbons is to quantify the 'source strength' represented by the free-phase hydrocarbons. The term 'source strength' is used to describe the intensity with which dissolved chemicals are released from a hydrocarbon into water. Source strength can be expressed as a mass per time per unit contact area:

$$S = K_{\mathrm{m}}A$$

where S is source strength (mass per time per meter squared), K_{m} is mass exchange coefficient (mg per meter squared per second), A is contact area (meter squared).

Table 4 Solubility of crude oil with various API gravities (milligrams/kilogram soil)

	Distilled water		
API gravity	*25°C*	*35°C*	*45°C*
11	3.5	4.5	6.0
15	4.0	5.0	7.0
20	10.0	14.0	18.0
28	65.0	75.0	90.0

The variable A is the interface across which mass exchange occurs over a given period of time. Due to the variety of hydrocarbon distribution patterns in the soil, estimating A is usually qualitatively defined. Maximum mass-exchange coefficients estimated by the US Geological Survey in 1984 were reported to be: 1.0 $\mathrm{mg\,m^{-2}\,s}$ for gasoline and tar oil; 0.01 for fuel oil, diesel, and kerosene; and 0.001 for lubricating oils and heavy fuel oil.

Petroleum Hydrocarbon Transport in Groundwater

Hydrocarbon transport in groundwater includes both the movement of the free phase in contact with the capillary fringe or within the groundwater, and the dissolved phase within the saturated zone. The movement of both plumes is controlled not only by the physical characteristics of the hydrocarbon as it enters the groundwater, but also by the media through which it flows.

Free-Phase Groundwater Plume

The free-phase component of the contaminant plume normally resides in greatest thickness at the point at which the hydrocarbon enters the groundwater. The lateral movement of the free-phase plume is controlled by those soil properties that control the movement of fluids and by the interaction of the free-phase plume with the groundwater gradient.

A situation can arise in which a free-phase hydrocarbon with a density less than water (an NAPL such as benzene) is found below the water table. One mechanism that accounts for this situation is the result of a water table increase and the entrainment of the benzene below or within a low-permeability sediment.

A dissolved phase can subsequently originate from the trapped free-phase liquid under the low-permeability layer. The rate at which the dissolved phase is created and its concentration can be estimated from solubility data and from information on the chemical concentration of the compound.

Dissolved-Phase Groundwater Plume

Most hydrocarbon fractions are immiscible with water. Some components in hydrocarbons, however, are soluble (e.g., BTEX), as are additives such as MTBE. This property provides a useful means for characterizing hydrocarbons and for understanding whether BTEX concentrations represent near-saturation conditions. Once the soluble fraction of the hydrocarbon has entered the groundwater, its movement is controlled by the same advective and dispersive processes used for estimating the mass transport of groundwater.

Length of Petroleum Hydrocarbon Plumes

Recent examination of hydrocarbon plumes has resulted in the finding that plumes rarely exceed 80 m in length. These observations in turn provide evidence for consideration and/or adoption of natural attenuation strategies as a remedial approach in documents such as the American Society for Testing and Materials *Standard Guide for Remediation by Natural Attenuation at Petroleum Release Sites*.

Lawrence Livermore National Laboratory conducted a study of hydrocarbon plumes in groundwater for the California State Water Resources Control Board to support revision of the Leaking Underground Fuel Tank corrective action process. As part of the data interpretation, the lengths of the hydrocarbon plumes were summarized; these results are similar to the findings of a study by Chevron of 119 sites in California. The results of both studies and the number of sites examined are summarized in Table 5.

These studies also classified hydrocarbon plumes in terms of the change in plume geometry for benzene. The plumes were described as 'shrinking,' 'no significant trend,' and 'increasing.' The results of this classification as percentages of the total sites are shown in Table 6.

Analytical solutions for groundwater transport and attenuation of hydrocarbons used to estimate the length of hydrocarbon plumes have been found to be similar to the Lawrence Livermore and Chevron experimental findings. The similarities between modeled and observed plume lengths include the finding that BTEX plumes may be relatively short (<90 m). Longer plumes might be expected when groundwater velocity is high (>3 cm day^{-1}) and when attenuation rates are relatively low. Modeling results of hydrocarbon plumes also suggests that BTEX plumes may reach stable or steady state conditions within 3–7 years.

Table 5 Results of two studies on lengths of hydrocarbon plumes in groundwater

Plume length (L)	*Lawrence Livermore (271)*	*Chevron (62)*
$L < 30$ m	50%	47%
30 m $< L <$ 80 m	40%	40%
$L > 80$ m	10%	13%

Table 6 Results of two studies on changes in hydrocarbon plume geometry

	Lawrence Livermore (271)	*Chevron (62)*
Shrinking	59%	51%
No significant trend	33%	40%
Increasing	8%	9%

Transport in Fractures

Fractures such as joints, faults, or bedding-plane separations can occur in various geologic strata. The significance of a release of hydrocarbons into a fracture system is (a) the velocities can be greater than for an unconsolidated media, and (b) the distribution of the hydrocarbons within the fracture system becomes difficult to define.

The flow of hydrocarbons in a fracture network is controlled by a number of factors, including the fracture aperture (i.e., the distance between the fracture walls), fracture density (i.e., spacing between the fractures), and fracture length and orientation. The inability to easily measure fracture apertures is primarily due to their size (10–100 μm) and the interferences caused from drilling or excavation stress-relief. As a result, aperture values are usually estimated using the 'cubic law.' The cubic law is based on the relationship between hydraulic gradient and the discharge for a viscous fluid flowing between two parallel smooth surfaces. In reality, apertures can vary greatly in size and wall roughness; these are not easily measured.

The transport of a dissolved-phase hydrocarbon in a fractured porous medium occurs by convection through the fractures, advection into the fracture wall matrix, and diffusion into the fracture walls (assuming that the material is porous). In most cases, advection into the wall is negligible; diffusion controls solute transport from a rapidly moving liquid through the fractures into the relatively immobile pore water of the wall matrix.

The process of matrix diffusion can theoretically retard the overall migration of a hydrocarbon by several orders of magnitude relative to transport within the fracture system. For example, modeling simulations show that, in a scenario, that included matrix diffusion, chloride migrating through a 25-μm aperture fracture in a glacial clay would take about 50 years to reach the underlying aquifer present at a depth of 10 m. For the same modeling scenario, but with no matrix diffusion, the solute would reach the aquifer within about half a day.

For NAPLs moving vertically through a fracture system, the transport is rapid after the release; diffusion has little impact as the aperture size and hydrocarbon properties (density, interfacial tension, and viscosity) impacts the rate of flow. Once the hydrocarbon becomes immobilized within the fracture system, diffusion into the porous matrix occurs. The diffusion of the hydrocarbon into the porous matrix significantly complicates the remediation of the hydrocarbons present within the porous media.

Remediation of Petroleum Hydrocarbon-Impacted Soil and Groundwater

Physical Separation

There are three significant physical separation processes that are utilized for petroleum hydrocarbon-impacted soil. Each of these remedial technologies uses equilibrium driving forces as the primary separation mechanism.

Soil vapor extraction Soil vapor extraction (SVE) is a vacuum extraction process. It utilizes the volatility of the petroleum hydrocarbons as the primary method of remediation. SVE works by utilizing vapor pressure differences between petroleum and water or petroleum and soil. The process is generally utilized to remove the more volatile petroleum products such as gasoline and aviation fuel. It can also be used to reduce the mobility and hazard of diesel, kerosene, and some heating oils by removing the more volatile fractions from the soil. SVE is usually limited to chemicals with a boiling point less than 250°C and high vapor pressure (greater than 0.5 mmHg). Chemicals with a Henry Law constant greater than 100 atmospheres tend to be removable from water using SVE.

During the SVE process, slotted wells are placed in the vadose zone in or around the contaminated soil. The well is sealed at the surface and connected to a vacuum pump. A vacuum is applied to the slotted wells, which causes a subsurface airflow to pass over the adsorbed soil–petroleum matrix. Utilizing vapor-phase equilibrium as the driving force, volatile components migrate from the liquid phase into the vapor phase and then into the slotted wells. Once these components reach the slotted wells, they are removed via the vacuum extraction system.

If the soil is initially wet, it is usually desiccated before there is significant recovery of the petroleum constituents. Low soil permeability and high soil-moisture content inhibit airflow and therefore inhibit petroleum constituent recovery. In general, the efficacy of SVE is enhanced whenever the temperature of the constituents or soil can be increased. Techniques that provide this increased temperature include microwaves, heated and injected air, and injection of steam.

Soil-washing (solvent extraction) Soil-washing is a solvent-extraction process. It uses the difference in the solubility of petroleum and the soil in water as the driving force. First, contaminated soils are usually excavated (although *in situ* soil flushing systems do exist) and the soil screened to remove large rocks and materials. If there are large amounts of clay or other very fine soil particles, soil-washing can be limited by the ability of the soil particle to settle after the washing/extraction has been accomplished.

Second, a surfactant is typically added to the petroleum-contaminated soil. Early surfactants were simple detergents. These detergents work by attaching a hydrophobic part of the detergent molecule to the petroleum while the hydrophilic part of the detergent molecule associates with water molecules. This association makes the petroleum-contaminated soil particle soluble in water.

When this principle is applied to soil-washing, the use of detergent-type surfactants creates an oil–water emulsion that must be treated, and the suspended soil must be separated from the solution. A major impediment to using soil-washing as a remediation technique has been the inability of this process to remediate fine-particle substrates and to treat the wash water in a cost-effective way.

Newer surfactants have been developed that form a separate phase that is easily separated from the soil-water phase. If there are many fine (e.g., clay) particles in the soil, mechanical separation of the soil from the liquid phase may be instituted in order to remove the water from the soil.

Some soil-washing units utilize extracting solvents such as triethylamine, carbon dioxide, or propane. For heavy oil recovery, water with an added surfactant is the most common processing solvent. Some new surfactants also contain many of the base nutrients necessary to support microbiological activity.

Thermal desorption Thermal desorption is a process that heats contaminated soils with an external heat source such as heated air flow or with an internal heat source such as an oil-heated screw feeder. The major difference between thermal desorption and incineration is that thermal desorption either captures or combusts volatilized chemicals emanating from the soil.

Thermal desorbers utilize several different designs. Rotary drum driers utilize either co-current or countercurrent heated airflow. Thermal screw designs use hollow screw feeders that usually contain hot oil as the heat source; the thermal screw design has one inherent advantage, that it can maintain a low-oxygen atmosphere in the chamber. Conveyor furnaces move the soil through a heated chamber on a metal belt. It is similar to the rotary drier in principle.

Thermal desorption systems usually operate at 200°–300°C, but some are operated at temperatures as high as 600°C. In order to determine the temperature at which a thermal desorber is to be operated, the boiling point range of the treated petroleum mixtures should be considered. Gasoline volatilizes at 40°–225°C, kerosene at 180°–300°C, diesel at 200°–340°C; lubricating oil and crude oil have constituents that volatilize at temperatures greater than 400°C. However, the amount of energy that must be added to the desorber to volatilize the hydrocarbons in the soils is also dependent in many systems on soil particle size and moisture content.

Biochemical Reactions

Several remediation technologies are built around biodegradation processes in which, ideally, the final products are CO_2 and water. The following subsections discuss common remediation techniques that utilize biochemical reactions. It is important to note that these techniques, like every successfully engineered bioremediation process, must create an environment where there are sufficient oxygen and nutrients and where the buildup of waste products such as carbon dioxide, water, and dead cells can be controlled.

Natural attenuation and biopiles Natural attenuation utilizes the natural soil environment to decontaminate a petroleum hydrocarbon-impacted soil. It is popular, because it is one of the most cost-efficient technologies available. However, it may not be the most time-effective approach available.

'Biopiles' refers to the controlled biodegradation of the organic fraction of a solid waste pile and is very similar to compost piles and sludge stabilization piles. Piles of petroleum hydrocarbon-impacted soils are amended with nutrients and water to stimulate microbial activity and the subsequent decomposition of hydrocarbons. This approach has been utilized for centuries in Asia and is used in developed countries to degrade many different types of waste.

Monitored natural attenuation (MNA) and biopiles seem to have lost appeal since 1995 as a technology for remediating contaminated soils. Several barriers to biopile use include a lack of confidence in the technology, costs, lack of understanding by states, and extended cleanup time.

Land farming This treatment technology is commonly known as land treatment, land cultivation, land spreading, land disposal, land application, refuse farming, and sludge farming. The primary objective of land farming is the degradation of a chemical or waste residing in topsoil by soil microorganisms. During land farming, hydrocarbon-impacted soil amended with nutrients and/or microorganisms is spread in thin layers on a field. Over time, microbial processes degrade the hydrocarbons.

Landfills Despite pollution often associated with unlined landfills, those which are lined and well managed are basically huge biopiles supporting the same biochemical processes that occur in small-scale biopiles and land farming.

Physicochemical Reactions

Some remediation techniques utilize a combination of physical as well as chemical reactions to remediate soils, thereby earning the name 'physicochemical reaction processes.' Some of the more common techniques are discussed below.

Solidification During the solidification remediation technique, a solidifying agent such as Portland cement or a calcium-rich pozzolanic material such as coal fly ash is mixed with a contaminated soil. A chemical process creates a new, hardened solid encapsulating the soil and rendering petroleum hydrocarbons immobile. Solidification is usually chosen when it is necessary to decrease the leachability (i.e., mobility) of the contaminants on the soil.

Asphalt incorporation Asphalt incorporation has been used for many years to recycle petroleum-contaminated soils. Contaminated soil is transported to a hot-batch asphalt plant. The volatile chemicals are vaporized in the initial combustion chamber. The lighter ends are volatilized in the asphalt dryer (200°–250°C) and are subsequently combusted. The heavy ends remain in the asphalt. The soil contaminated with the heavy ends makes a very good aggregate replacement for clean fine aggregate if the soil particle size is appropriate.

The production of asphalt from the contaminated soil requires a complete investigation of the properties of the petroleum-contaminated soil, the percentage of petroleum in the mix, and the quality of the petroleum. A special treatment and testing program may be needed to produce and verify a quality product meeting acceptable standards and specifications, such as ASTM specifications.

The production of asphalt concrete with petroleum-contaminated soil is considered a practical and economical procedure that produces a useful and marketable product inexpensively, and it cleans the environment. The final product is used as a construction material for secondary roads, roadbeds, road subbase, impermeable layers for landfill and containment facilities, or stabilizers for steep embankments.

Incineration Contaminated soil is burned in a suitable incinerator at temperatures of 600°–1000°C. Air emissions are generally required to be captured and treated if the soil contains significant quantities of chlorine or volatile metals such as mercury.

Vitrification Vitrification uses large quantities of electrical energy to physically melt the contaminated soil and to oxidize any petroleum hydrocarbons in the soil. The resulting melt vitrifies into a glass-like mineral that encapsulates any metals and other chemicals contained in the melt for hundreds or thousands of years.

Summary

Contamination of soils and of groundwater by hydrocarbons appears to be a more widespread and complex problem than previously recognized. The presence of various additives in petroleum products (including lead, Pb, as well as certain persistent organic compounds) further exacerbates the problem. Better techniques are needed for identifying, delineating, and remediating the problem in sites where it already exists (as a consequence of past practices or of mishaps such as oil spills or leaky fuel tanks), and – even more importantly – for prevention of the problem from occuring at new sites in the future.

See also: **Pollutants:** Biodegradation; Effects on Microorganisms; Persistent Organic (POPs); **Remediation of Polluted Soils**

Further Reading

American Petroleum Institute (1972) The migration of petroleum products in soil and groundwater. *Principles and Counter Measures*. Washington, DC: API.

American Society for Testing and Materials (1996) *Standard Guide for Remediation by Natural Attenuation at Petroleum Release Sites*. (Draft.) ASTM.

Bruce L and Schmidt G (1994) Hydrocarbon fingerprinting for application in forensic geology: review with case studies. *American Association of Petroleum Geologists Bulletin* 78(11): 1692–1710.

Calabrese EJ and Kostecki PT (1992) *Risk Assessment and Environmental Fate Methodologies*. Chelsea, MI: Lewis Publishers.

Coleman W, Munch J, Stricher R, Ringhand H, and Kopfler F (1984) The identification and measurement of components in gasoline, kerosene, and No. 2 fuel oil that partition into the aqueous phase after mixing. *Archives of Environmental Contamination and Toxicology* 13: 171–178.

Domenico P (1987) An analytical model for multidimensional transport of a decaying contaminant species. *Journal of Hydrology* 91: 49–58.

Dragun J (1991) Geochemistry and soil chemistry reactions occurring during *in situ* vitrification. *Journal of Hazardous Materials* 26: 343–364.

Dragun J (1998) *The Soil Chemistry of Hazardous Materials*, 2nd edn. Amherst, MA: Amherst Scientific Publishers.

Hillel D (1988) Chapter 7 – Movement and Retention of Organics in Soil: A Review and a Critique of Modeling. *Petroleum Contaminated Soils. vol. I. Remediation Techniques, Environmental Fates, Risk Assessment*. In: Kostecki P and Calabrese E (Eds) Chelsea, MI: Lewis Publishers.

Johnson P, Kemblowski M, and Colthart J (1990) Quantitative analysis for the cleanup of hydrocarbon-contaminated soils by *in-situ* soil venting. *Ground Water* 28: 413–429.

Kaplan I, Galperin Y, Lu S, and Lee R (1997) Forensic environmental geochemistry: differentiation of fuel-types, their sources and release time. *Organic Geochemistry* 27(5/6): 289–317.

Kueper B and McWhorter D (1991) The behavior of dense, non-aqueous phase liquids in fractured clay and rock. *Ground Water* 29: 716–728.

Kung S (1990) Preferential flow in a sandy vadose zone: field observation. *Geoderma* 46: 51–58.

Kung S (1990) Preferential flow in a sandy vadose zone. 1. mechanism and implications. *Geoderma* 46: 59–71.

Morrison R (2000) Application of environmental forensic techniques for age dating and source identification of a contaminant release. *International Journal of Environmental Forensics* 1(2): 131–153.

Murphy B and Morrison R (2002) *Introduction to Environmental Forensics*. Oxford, UK: Academic Press/Elsevier Science.

Ryan R and Dhir V (1993) The effect of soil-particle size on hydrocarbon entrapment near a dynamic water table. *Journal of Soil Contamination* 2(1): 59–92.

Hydrodynamic Dispersion *See* Solute Transport

HYDRODYNAMICS IN SOILS

T P A Ferré and A W Warrick, University of Arizona, Tucson, AZ, USA

Introduction

'Hydrodynamics' means 'the movement of water.' In contrast to hydrostatics, hydrodynamics considers the response of a fluid to an imposed energy gradient. The subfield of soil-water dynamics can be further subdivided into flow through a saturated or unsaturated medium, and steady-state or transient processes. Consideration of a saturated medium applies to groundwater flow problems. Flow through the vadose zone requires study of saturated and unsaturated flow through variably saturated media. Both saturated and unsaturated flow problems may be steady-state or transient. Soil-water dynamics is so central to topics involving the movement of water, heat, and solutes through the subsurface that it is fundamental to most, if not all, soil physical processes and groundwater flow problems.

Steady-State Flow Through a Saturated Medium

The Darcy equation describes the response of a fluid in a porous medium to an imposed energy gradient and forms the foundation of discussions of hydrodynamics. The one-dimensional form of the Darcy equation can be written as:

$$Q = -A\frac{k\rho g}{\mu}\frac{\Delta(z+\psi)}{\Delta L} \qquad [1]$$

where Q is the flow (L^3/T), A is the cross-sectional area perpendicular to flow (L^2), k is the intrinsic permeability (L^2), ρ is the fluid density (M/L^3), g is the acceleration due to gravity (L/T^2), μ is the dynamic viscosity of the fluid (M/LT), ψ is the pressure head (L), z is the elevation (L), and L is the distance in the direction of flow (L). The negative sign is necessary due to the definition of a gradient to define flow as occurring from a region of higher energy to lower energy. This can be rewritten in terms of the flux density, $q(L/T)$ as:

$$q = \frac{k\rho g}{\mu}\frac{\Delta(z+\psi)}{\Delta L} \qquad [2]$$

Flow occurs in the direction of decreasing energy of the fluid, and the rate of flow depends directly upon the ability of the medium to permit flow (permeability), the fluid density, and the acceleration due to gravity, and inversely upon the resistance of the fluid to flow (dynamic viscosity). Introducing the hydraulic conductivity, K (L/T) allows simplification to:

$$q = -K\frac{\Delta(z+\psi)}{\Delta L} \qquad [3]$$

For flow in the horizontal, x-direction, the flux density equals:

$$q_x = -K\frac{\Delta\psi}{\Delta x} \qquad [4]$$

The flux density for flow in the vertical, z-direction, is:

$$q_z = -K\left(1+\frac{\Delta\psi}{\Delta z}\right) \qquad [5]$$

Furthermore, if the medium is homogeneous and flow is steady-state, then the gradients of the hydraulic and pressure heads must be constant during one-dimensional flow. That is, the pressure head and hydraulic head will vary linearly with distance in the direction of flow. This is shown for a 200-cm-long, homogeneous horizontal column for which the left boundary is maintained at a constant pressure head of 70 cm and the right boundary is held at a constant pressure head of 0 cm (Figure 1).

The two-dimensional form of the Darcy equation follows directly from the one-dimensional form. For instance, two-dimensional flow in the x-z plane is described as:

$$q_i = -K_{i,j}\frac{\Delta(z+\psi)}{\Delta x_j} \qquad [6]$$

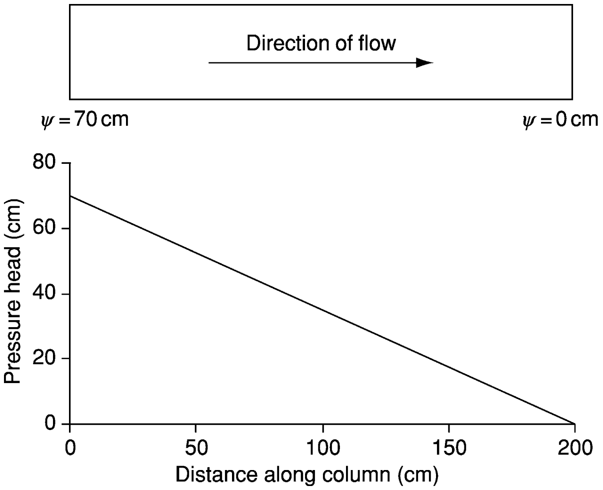


Figure 1 Pressure-head profiles along a horizontal column during steady-state saturated flow.

The subscripts i and j represent either x or z in this case. The hydraulic conductivity matrix, K_{ij}, allows for flow in the x-direction due to gradients in the x-direction ($K_{x,x}$) and gradients in the z-direction ($K_{x,z}$). Furthermore, the convention is followed that a sum is taken over the index that is included twice on the right-hand side of the equation. For example the flux density in the x-direction is:

$$q_x = -K_{x,x}\frac{\Delta(z+\psi)}{\Delta x} - K_{x,z}\frac{\Delta(z+\psi)}{\Delta z} \qquad [7]$$

The flux expression can be written in terms of partial derivatives as:

$$q_i = -K_{i,j}\frac{\partial(z+\psi)}{\partial x_j} \qquad [8]$$

While the modifications of the equations to describe multidimensional flow are straightforward, the results are not trivial. Specifically, for one-dimensional geometries, flow is constrained to a single direction. For multidimensional cases, water will tend to flow preferentially in the direction of highest permeability. This phenomenon is referred to as 'anisotropic flow' if it arises due to a directional dependence of the hydraulic conductivity (e.g., along or across bedding planes).

Transient Flow Through a Saturated Medium

If the energy gradient applied to the fluid in a medium is constant in time, the flow through the medium will achieve a steady-state condition given enough time. When this state is reached, the total volume of water entering any region of the porous medium in a period of time will be exactly equal to the total volume of water leaving that region. The region can be defined as a point for one-dimensional flow or as an area or a volume for two- and three-dimensional flow, respectively. Before steady state is reached, the medium responds in a transient manner to the changes in the imposed energy gradient. The mechanism by which the medium responds is through a change in the volume of water stored within the medium.

Consider the three-dimensional region shown in Figure 2. Each edge has a constant length, ΔL. Flow is two-dimensional in the x and z directions. At steady state the sum of q_x in and q_z in will be equal and opposite to the sum of q_x out and q_z out. However, under transient flow conditions, these sums will not be equal. The net difference in the inflow and the outflow will equal the change in the volume of water stored, ΔV_w, per unit time elapsed, Δt, per unit area of the region. This is written as a mass balance equation:

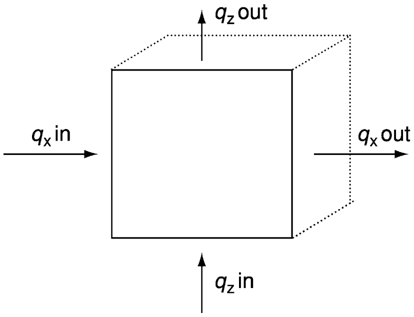


Figure 2 Three-dimensional unit cell for consideration of conservation of mass. The flux density into (q in) and out of (q out) the cell is shown in the x and z directions.

$$(q_x \text{ in} + q_z \text{ in})(\Delta L\Delta L) - (q_x \text{ out} + q_z \text{ out})(\Delta L\Delta L) = \frac{\Delta V_w}{\Delta t} \qquad [9]$$

For the fluxes shown in Figure 2, the net flux in the positive x direction can be defined as $\Delta q_x = q_x \text{ out} - q_x \text{ in}$. Substituting for the net fluxes and dividing by the areas of the sides gives:

$$-\Delta q_x - \Delta q_z = \frac{\Delta V_w}{\Delta t\Delta L^2} \qquad [10]$$

Dividing both sides by $-\Delta L$ gives:

$$\frac{\Delta q_x}{\Delta L} + \frac{\Delta q_z}{\Delta L} = -\frac{\Delta V_w}{\Delta t\Delta L^3} \qquad [11]$$

By considering infinitesimally small regions and infinitesimally small time steps, ΔL and Δt approach zero and the mass balance equation can be written in partial differential form:

$$\frac{\partial q_i}{\partial x_i} = -\frac{\partial}{\partial t}\left(\frac{\Delta V_w}{\Delta L^3}\right) \qquad [12]$$

where x_i is x and z. Combined with the Darcy equation, this gives:

$$\frac{\partial}{\partial x_i}\left(K_{i,j}\frac{\partial(z+\psi)}{\partial x_j}\right) = \frac{\partial}{\partial t}\left(\frac{\Delta V_w}{\Delta L^3}\right) \qquad [13]$$

This can be extended to three dimensions by allowing i to correspond to x, y, and z.

Saturated water flow is caused, in part, by changes in the pressure head of the fluid. Therefore, it is convenient to describe the changes in the volume of stored water per unit area in terms of the response of the medium to changes in pressure head as well. The specific storage, S_s, is defined as the volume of water released from storage per unit volume of medium per unit decrease in the pressure head of the water:

$$S_s = \left(\frac{\Delta V_w}{\Delta \psi \Delta L^3}\right) \qquad [14]$$

The specific storage is a lumped term that describes the results of two physical phenomena: medium compaction and water expansion. Water that resides in a medium and is under a positive pressure head acts to expand the matrix of the medium. If the water pressure is reduced, the effective stress on the medium will increase and the medium will compact. By compacting, the porosity of the medium is reduced as water is released from storage. Simultaneously, the decreased pressure of the water allows the water itself to expand, thereby occupying more volume per unit mass of water. This is also accompanied by a release of water from storage. Of the two mechanisms, the first is dominant due to the very low compressibility of water.

Substituting the specific storage into the mass balance equation gives:

$$\frac{\partial}{\partial x_i}\left(K_{i,j}\frac{\partial \psi}{\partial x_j}\right) = S_s \frac{\partial \psi}{\partial t} \qquad [15]$$

This form of the mass balance equation conveniently describes the transient response of a porous medium to changes in the applied pressure heads. Physically, this demonstrates that a change in pressure at some location within a medium will take a finite time to propagate to a distant point in the medium. This delay is caused by changes in water storage (either increases in response to an increased pressure head or decreases due to a decreased pressure head) that are caused at each point past which the pressure change is transmitted. The time that it will take for the pressure pulse to be transmitted increases with an increase in the value of the specific storage of the medium. This time also increases with a decrease in the hydraulic conductivity of the medium, because this leads to slower flow of water through the medium in response to an imposed energy gradient.

Steady-State Flow Through an Unsaturated Medium

The Darcy equation is equally applicable to steady-state flow through both saturated and unsaturated media. However, there is a critical difference from the preceding analysis when this relationship is applied to unsaturated flow. Specifically, the hydraulic conductivity (or intrinsic permeability) of the medium is no longer independent of the pressure head of the water in the medium. As the pressure head of the water decreases below the air entry pressure of the medium, the volumetric water content of the medium (θ), defined as the volume of water per unit volume of medium, also decreases. As a result, some of the pore spaces are no longer water-filled and, as such, are not available for water flow. This reduction in the water-filled porosity of the medium reduces the ability of the medium to transmit water under an implied energy gradient. This can be described as a decrease in the hydraulic conductivity with decreasing water content. The equation describing unsaturated flow is deceptively simple:

$$q_i = -K_{i,j}(\theta)\frac{\partial(z+\psi)}{\partial x_j} \qquad [16]$$

The ramifications of the dependence of the hydraulic conductivity on the volumetric water content can be quite dramatic. For example, consider the case of steady-state, one-dimensional, horizontal flow through a homogeneous medium. The pressure-head profile is shown in Figure 3 for a 200-cm-long, sand-filled column with the left boundary held at a constant pressure head of 0 cm and the right boundary held at a constant pressure head of −70 cm. Notice that, unlike the case of steady-state one-dimensional flow through a saturated medium, the pressure head varies nonlinearly in the direction of flow. This occurs because the hydraulic conductivity decreases steadily with decreasing pressure head from the saturated value at the left boundary to a much lower value at the right boundary. To maintain steady-state flow throughout the column, the magnitude of the gradient must increase to compensate exactly for this decreasing hydraulic conductivity. Therein lies the difficulty of calculating the pressure-head distribution during unsaturated flow. Even under the apparently simple conditions of steady-state, one-dimensional flow, the pressure-head distribution depends on the change in the hydraulic-head

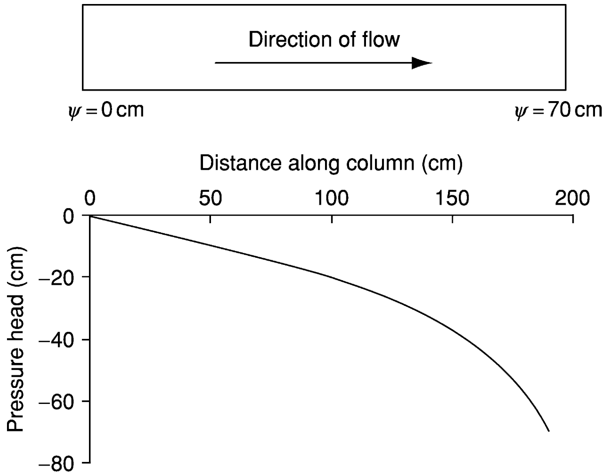


Figure 3 Pressure-head profiles along a horizontal column during steady-state unsaturated flow.

gradient through the system. The gradient depends upon the hydraulic conductivity distribution. However, the hydraulic conductivity distribution depends upon the water content distribution, which, in turn, depends upon the pressure-head distribution. Equations that describe such a circular dependence of the hydraulic parameters on the energy state and of the energy state on the hydraulic parameters are nonlinear.

An alternative approach to describing unsaturated flow in the absence of gravity is to consider the matric flux potential, ϕ, which is defined as:

$$\phi = \int_{-\infty}^{\psi} K(h)\mathrm{d}h \qquad [17]$$

The derivative of the matric flux potential is related to the flux as:

$$-\frac{\mathrm{d}\phi}{\mathrm{d}x} = q_{\mathrm{x}} \qquad [18]$$

Given that the flux is constant throughout the column under steady-state flow, the derivative of the matric flux potential will also be constant with x. That is, ϕ will vary linearly with x. The value of ϕ at any position will depend on the values of pressure head at the ends of the column and on the conductivity function $K(\psi)$.

Transient Flow Through an Unsaturated Medium

Transient unsaturated flow is fundamentally different than saturated flow and steady-state unsaturated flow. During transient unsaturated flow, water enters pores that were previously occupied by another fluid. Commonly this fluid is air; more commonly, it is assumed that the displacement of the resident air does not impede the advance of water into a pore. This is the underlying assumption of the Richards equation, a special expression of mass conservation that describes the movement of water through an unsaturated porous medium. The Richards equation can be written in many forms. One form that allows ready comparison with the mass balance equation developed for saturated flow is written:

$$\frac{\partial}{\partial x_{\mathrm{i}}}\left(K_{\mathrm{i,j}}(\theta)\frac{\partial(z+\psi)}{\partial x_{\mathrm{j}}}\right) = \frac{\partial\theta}{\partial t} \qquad [19]$$

Direct comparison with the mass balance equation developed for transient, saturated flow shows two differences. As for steady-state unsaturated flow, the hydraulic conductivity is now a function of the water content. This leads to the same nonlinear behavior seen above. In addition, the storage term is now written as the change in the unitless volumetric water content, θ, per unit time. The use of the change in water content alone to describe storage changes suggests that the mechanisms that allow for changes in water storage in a saturated medium do not apply in an unsaturated medium. In fact, medium compaction and water expansion do occur under conditions of negative pressure head in the water phase. However, the amount of storage change that can occur owing to changes in water content are far larger than those that can occur owing to both of these mechanisms combined. As a result, they can commonly be ignored for unsaturated conditions to allow for more convenient solutions.

Transient unsaturated flow differs from transient saturated flow as a result of the nonlinearity of the Richards equation, through the dependence of the hydraulic conductivity on the water content and of the large potential for water storage changes through changes in the volumetric water content. The first of these effects gives rise to relatively slow movement of water under a given energy gradient in an unsaturated medium compared with the response to the same gradient under saturated conditions. A secondary result is the formation of sharp wetting fronts. (*See* **Infiltration**.) The second of these effects results in even slower propagation of pressure changes through an unsaturated medium due to the large capacity for changes in water storage in response to pressure changes. The validity of the Richards equation and of pressure–saturation relationships determined under static conditions for describing transient flow was still a topic of research at the beginning of the twenty-first century.

See also: **Darcy's Law**; **Diffusion**; **Heat and Moisture Transport**; **Infiltration**

Further Reading

Darcy H (1856) *Les Fontaines Publique de la Ville de Dijon*. Paris, France: Dalmont.

Gardner WR (1958) Some steady state solutions of the unsaturated moisture flow equation with application to evaporation from a water table. *Soil Science* 85(4): 228–232.

Richards LA (1931) Capillary conduction of liquid through porous media. *Physics* 1: 318–333.

HYSTERESIS

J H Dane, Auburn University, Auburn, AL, USA
R J Lenhard, INEEL, Idaho Falls, ID, USA

Introduction

When studying fluid flow in porous media such as soils, we can analyze fluid behavior at the molecular, the microscopic, or the macroscopic level. At the microscopic level, we assume that only one phase is present at a given physical point *P* and at a given time, i.e., at *P* only the solid phase or one of the fluids is present. This is in contrast to the macroscopic level, where we assume that all phases can be present at the same physical point and at the same time (principle of superposition). At both the microscopic and the macroscopic level, all variables are considered to be continuous in space and time, so that they are differentiable. Fluids present in the soil are wetting fluids, nonwetting fluids, or of intermediate nature. Although this discussion can be extended to any set of fluids, for the sake of simplicity only the presence of water (wetting fluid) and/or air (nonwetting fluid) is considered here. When only water occupies the pore spaces, the soil is said to be completely saturated. When only air is present, the soil is said to be completely dry. In most cases, both air and water are present, and a relationship exists between the energy status of the soil water and the volumetric water content. The governing potential that determines the water content is the matric potential, ϕ_m (energy per unit volume, with units: joules per cubic meter, or newtons per meter squared, or pascals; energy per unit mass, with units joules per kilograms: or energy per unit weight, with units joules per newton or meters). It consists of three components that bind the water to the solid phase:

$$\phi_m = \phi_p + \phi_a + \phi_\omega \qquad [1]$$

where ϕ_p, ϕ_a, and ϕ_ω refer to the pressure potential, the adsorptive potential, and the osmotic potential, respectively, all at the microscopic level. The osmotic forces are due to concentration differences between the bulk solution and the diffuse double layer. The adsorptive forces are forces of attraction between the water molecules and the solid phase and operate at very short distances only. Assuming the osmotic forces to be part of the adsorptive forces, eqn [1] simplifies to:

$$\phi_m = \phi_p + \phi_a \qquad [2]$$

When the water is at static equilibrium, the total mechanical potential, ϕ, is:

$$\phi = \phi_g + \phi_m \qquad [3]$$

where the gravitational potential, ϕ_g, is constant when considering a horizontal cross-sectional area through a pore located at *P*. By assuming $\phi = 0$ at some reference level, it follows that $\phi = 0$ everywhere. Hence, the matric potential is constant through this horizontal cross section, i.e.,

$$\phi_m = -\phi_g \qquad [4]$$

When sufficient water is present, the value for the matric potential is completely determined by the value of the pressure potential at the water–air interface, i.e., the adsorptive and osmotic forces in the liquid at the interface are negligible. This will be the case as long as water is present so that a concave air–water interface exists. The qualitative potential distributions in a horizontal cross section of a pore are shown in Figure 1. The solid phase is at $x = 0$, while the concave air–water interface is at the dashed line. The question thus arises as to what determines

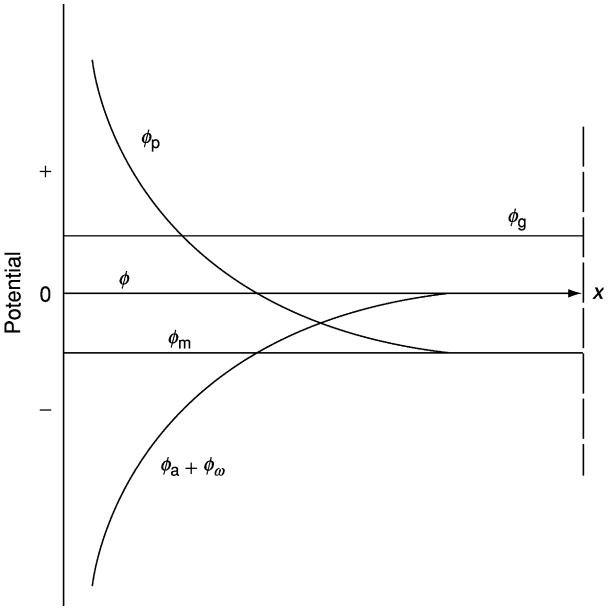


Figure 1 The qualitative potential distributions in a horizontal cross section of a pore inside a soil column with its water at static equilibrium. The pore water has an assumed concave meniscus. The matric potential $\phi_m = \phi_p + \phi_a + \phi_\omega$, where ϕ_p refers to the pressure potential, ϕ_a the adsorption potential, and ϕ_ω the osmotic potential, all at the microscopic level. The latter only exists inside the diffuse double layer. The total mechanical potential is given by $\phi = \phi_m + \phi_g$, where ϕ_g is the gravitational potential. ϕ is assumed to be zero at all elevations.

the pressure in the water just across the air–water interface.

The Capillary Pressure Concept

If more than one fluid is present, such as in an air–water system, the fluids are separated by an interface. If this interface is curved, a pressure jump, ΔP (newtons per meter squared), exists across the interface, which is determined by the Young and Laplace equation:

$$\Delta P = \sigma\left(\frac{1}{R_1} + \frac{1}{R_2}\right) \qquad [5]$$

where σ (newtons per meter) is the surface or interfacial tension, and R_1 and R_2 (meters) are the radii defining the curvature. For spherically shaped interfaces these radii of curvature are equal, and eqn [5] can be written as

$$\Delta P = \frac{2\sigma}{R} \qquad [6]$$

It is necessary to make conventions about the exact definitions of ΔP and R. ΔP can be defined:

$$\Delta P = P_l - P_g \qquad [7]$$

where the subscripts l and g refer to the liquid (water) and gas (air) phase, respectively. It should be noted, however, that the pressure jump across the interface can also be defined as the pressure in the gas phase minus the pressure in the liquid phase. Positive values can be assigned to R if R resides in the liquid phase, and negative values if it resides in the gas phase. It can thus easily be seen that $\Delta P > 0$ for a raindrop in air and $\Delta P < 0$ for water in a (hydrophilic) capillary tube or in a soil pore. In the case of a flat interface, $R \to \infty$ and $\Delta P = 0$. Hence no pressure jump exists across a flat interface.

As long as the air–water interface is concave toward the water, the matric potential in the pore water at a given point is determined by the pressure potential in the water at the air–water interface, which is determined by the water pressure at the air–water interface. When expressed on a mass basis, and assuming the air to be at atmospheric pressure ($P_g = 0$), the matric potential is defined:

$$(\phi_m)_{\text{mass}} = (\phi_p)_{\text{mass}} = \frac{P_l}{d_l} \qquad [8]$$

in joules per kilogram; while the matric potential on a volume basis is defined:

$$(\phi_m)_{\text{volume}} = (\phi_p)_{\text{volume}} = P_l \qquad [9]$$

in newtons per meter squared, or pascals; and the matric potential on a weight basis:

$$(\phi_m)_{\text{weight}} = (\phi_p)_{\text{weight}} = \frac{P_l}{d_l g} \qquad [10]$$

in meters, where d_l is the density of water (kilograms per cubic meter) and g is the gravitational field strength (newtons per kilogram). The matric potential on a weight basis is often referred to as the matric head, h_m. It should be noted that the pressure potential increases, while the adsorptive and osmotic potentials decrease, when approaching the solid phase. The water pressure at the air–water interface, P_l, and consequently the matric potential are determined by eqns [6] and [7]. It should be emphasized that, for capillary water (R in the gas phase), values for P_l, and hence for the matric potential, are negative. Using the alternative definition of ΔP (eqn [7]), i.e., $\Delta P = P_g - P_l$, and assuming $P_g = 0$, results in values of the capillary pressure $P_c = -P_l$ (pascals) or the capillary pressure head $h_c = -P_l/d_l g$ (in meters). The use of P_c or h_c is sometimes preferred, to deal with positive rather than negative values.

Water-Retention Curve

The relationship between matric potential and volumetric water content (θ) is referred to by many names, such as 'water-retention curve,' 'soil-water characteristic curve,' 'pressure head–water-content relationship,' 'water content–matric potential curve,' and 'capillary pressure–saturation curve.' Stepwise reduction of the matric potential of an initially completely saturated soil, i.e., starting at a matric potential value of zero, results in the initial drainage curve (IDC; **Figure 2**). When no more water can be drained from the soil, the residual water content (θ_r) has been reached. Reversing the process at this point, by increasing stepwise the matric potential until it reaches zero, produces the main wetting curve (MWC; **Figure 2**). The two curves are noticeably different, which is referred to as 'hysteresis.' Reversing the process again yields the main drainage curve (MDC; **Figure 2**). Reversal points can start at any point during drainage or wetting and will form scanning curves, which are always inside the loop formed by the MWC and MDC.

All variables in the previous sections have been assumed to occur at the microscopic level, or at least not at a level greater than the pore level. Most of the measurements are, however, obtained at the macroscopic level, as is the case with water-retention curves. Transitions can be made from the microscopic level to the macroscopic level, and it can be shown that, under most conditions, expressions for

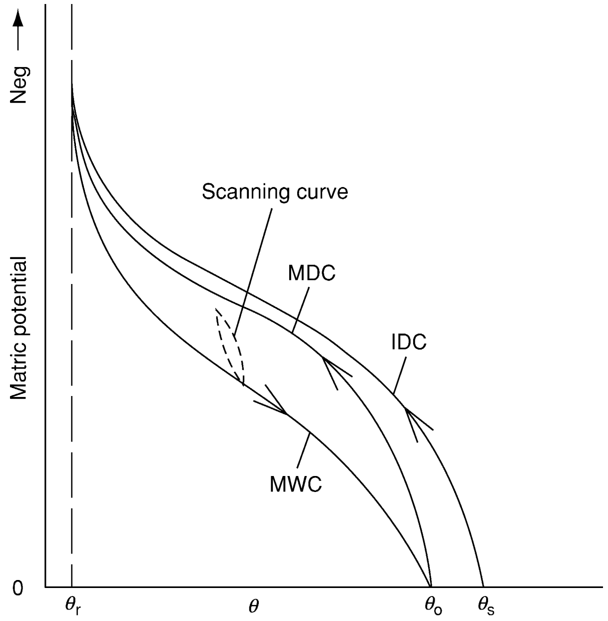


Figure 2 The initial drainage curve (IDC), the main wetting curve (MWC), the main drainage curve (MDC), and a scanning curve. θ refers to the volumetric water content and the subscripts r, o, and s, to residual, satiated, and saturated.

potential have the same form at both levels. Consequently, at the macroscopic level, average pressures within macroscopic volume elements are dealt with.

Causes of Hysteresis

Hysteresis in the water content–matric potential relation is attributed to: (1) the 'ink bottle' effect, (2) the contact angle effect, (3) air entrapment, and (4) deformation of the solid phase.

The ink bottle effect is probably the most important reason for hysteresis to occur and will be explained with the help of the Young and Laplace equation (eqn [6]). Envision a spherically shaped soil pore with two bottlenecks at opposite ends of the pore. The radius at the widest section of the pore is r_w (>0), whereas the radii of the bottlenecks, the smallest parts of the pore, are r_s (>0). One bottleneck is connected to a body of free water in which the pressure can be changed. The pore is small enough so that gravity plays no role in the drainage or imbibition of the pore water. Starting with a fully saturated pore, the water pressure in the body of free water is slowly reduced, and hence in the water in the pore, from zero to some negative value. The pressure in the water at which the pore will drain is determined by:

$$P_{l,d} = \Delta P = -\frac{2\sigma\cos\beta_d}{r_s} \qquad [11]$$

where R in eqn [6] has been replaced by $r_s/\cos\beta_d$ and β_d is the contact angle between the water and the solid phase during drainage. Once the pressure as expressed by eqn [11] has been reached, the pore will abruptly and completely drain. When the pressure in the water is then slowly increased, i.e., from some negative value in the direction of zero, part of the pore will slowly refill. The pore will suddenly completely fill when the pressure in the water has increased to:

$$P_{l,w} = \Delta P = -\frac{2\sigma\cos\beta_w}{r_w} \qquad [12]$$

where R has now been replaced by $r_w/\cos\beta_w$ and β_w is the contact angle during wetting. Assuming the contact angles during wetting and drying not to be too much different, it follows from eqns [11] and [12] that the pressure at which a pore drains is more negative than that at which it refills. Averaging this process over a number of pores, i.e., moving from the microscopic to the macroscopic level, results in data points that show that for the same water content the matric potential will be more negative during drainage than during wetting.

It has been shown that the contact angle between water and the solid phase is greater during imbibition of water than during drainage, i.e., $\beta_w > \beta_d$. According to eqns [11] and [12], this makes the pressure during drainage more negative than during wetting at the same water content, which reinforces the ink bottle effect.

Considerable effort needs to be applied to saturate a soil sample completely. This can be accomplished by first flushing the sample with CO_2, which readily dissolves in water, or by saturating the sample under vacuum. In this manner the IDC (**Figure 2**) can be established. Upon rewetting the soil, however, air cannot always escape and becomes trapped. This causes the water content to be smaller during wetting than during drying at the same value of the matric potential (same radius of curvature at the air–water interface). Because of entrapped air, the soil water content at zero value of the matric potential is often approximately 0.85 times the completely saturated value. The term 'satiated water content' is sometimes used in this respect; and the term 'apparent saturation' is used to account for hysteresis that is due to entrapped air or other nonwetting fluids.

Finally, swelling, shrinking, and aging phenomena may result in different soil structures during wetting and drying. One may argue, however, that the causes for hysteresis should strictly be explained by assuming the porous medium to be rigid. Changes in the soil structure obviously change the water-retention curve, as demonstrated when solutions of differing composition and concentration are used.

Modeling

It is obvious from Figure 2 that considerable differences may occur in water content values during drying and wetting if compared at the same value of the matric potential. As the ultimate goal is to predict water content values in the field by solving the general flow or the Richards equation for a set of initial and boundary conditions, it is important to account for hysteresis. Although the model of Dane and Wierenga shows this importance, the generated scanning curves do not form closed loops, and as such are not necessarily unique. The best available hysteresis model for air–water relations is probably Lenhard's.

The model by Lenhard *et al.* is the air–water version of the two- and three-fluid model proposed by Parker and Lenhard. The concept of apparent saturation ($\bar{\bar{S}}$) is used to model the effects of hysteresis on subsurface fluid behavior. Apparent saturation applies to fluids that are more wetting than the nonwetting fluid. Hence, for air–water fluid systems in porous media, apparent saturation pertains to water but not to air. The apparent water saturation ($\bar{\bar{S}}_w$) of an air–water system is defined as the summation of the effective water saturation ($\bar{S}_w$), defined:

$$\bar{S}_w = \frac{S_w - S_r}{1 - S_r} \qquad [13]$$

and the effective entrapped-air saturation ($\bar{S}_{at}$), defined:

$$\bar{S}_{at} = \frac{S_{at}}{1 - S_r} \qquad [14]$$

Consequently:

$$\bar{\bar{S}}_w = \frac{S_w + S_{at} - S_r}{1 - S_r} \qquad [15]$$

where $S_w = \theta/\epsilon$ is the water saturation, $S_r = \theta_r/\epsilon$ is the residual saturation, and $S_{at} = \theta_{at}/\epsilon$ is the entrapped air saturation; θ is the volumetric water content, θ_r is the residual water content, θ_{at} is the volumetric entrapped air content, and ϵ is the porosity.

To describe relations between fluid content and capillary pressure that account for hysteresis, two water-rentention curves are parameterized: one being the IDC (Figure 2) and the other being the MWC (Figure 2). It should be noted that these two curves (Figure 2) first need to be converted from θ (P_c) to $\bar{\bar{S}}_w$ (P_c) relations. For the IDC this poses no problem, because no entrapped air exists. Therefore, the apparent saturation is equal to the effective saturation. The IDC is described by the van Genuchten retention function:

$${}^{D}\bar{\bar{S}}_w = [1 + (\alpha_d P_c)^n]^{-m} \qquad [16]$$

where ${}^{D}\bar{\bar{S}}_w$ is the apparent water saturation of the IDC, α_d is a curve-fitting parameter for the IDC possessing the same units as $1/P_c$, P_c is the capillary pressure, n is a fitted shape parameter, and $m = 1 - 1/n$. The MWC is similarly described:

$${}^{I}\bar{\bar{S}}_w = [1 + (\alpha_w P_c)^n]^{-m} \qquad [17]$$

where ${}^{I}\bar{\bar{S}}_w$ is the apparent water saturation of the MWC (*I* is 'imbibition' or 'wetting') and α_w is a curve-fitting parameter for the MWC possessing the same units as $1/P_c$. The fitted parameter n is identical for the IDC and the MWC. The starting point for the MWC is the residual water content, while the end point is at a capillary pressure of zero. The conversion from ${}^{I}\bar{\bar{S}}_w$ to θ is accomplished by determining $\bar{S}_{at}$ at the ${}^{I}\bar{\bar{S}}_w$-value of interest. Once $\bar{S}_{at}$ is determined, then θ can be calculated from eqn [15]. Lenhard *et al.* modeled $\bar{S}_{at}$ to vary linearly from $\bar{S}_{at} = 0$ at the saturation-path reversal from the IDC (${}^{\Delta}\bar{S}_w$) to a maximum value at ${}^{I}\bar{\bar{S}}_w = 1$. Because the MWC is under consideration, the reversal point is at θ_r and $\bar{\bar{S}}_w = \bar{S}_w = {}^{\Delta}\bar{S}_w = 0$, i.e., there is no entrapped air at this point. When reversing from the IDC at any point other than θ_r, ${}^{\Delta}\bar{S}_w$ will be different from zero. The maximum value of $\bar{S}_{at}$ is a function of ${}^{\Delta}\bar{S}_w$ and is determined using the method of Land:

$$\bar{S}_{ar} = \frac{1 - {}^{\Delta}\bar{S}_w}{1 + R_{aw}(1 - {}^{\Delta}\bar{S}_w)} \qquad [18]$$

in which

$$R_{aw} = \frac{1}{{}^{I}\bar{S}_{ar}} - 1 \qquad [19]$$

where $\bar{S}_{ar}$ is the maximum entrapped air saturation for a given wetting path and ${}^{I}\bar{S}_{ar}$ is the maximum entrapped air saturation of the MWC. For an intermediate point along any wetting path, $\bar{S}_{at}$ is determined from:

$$\bar{S}_{at} = \bar{S}_{ar}\left(\frac{\bar{\bar{S}}_w - {}^{\Delta}\bar{S}_w}{1 - {}^{\Delta}\bar{S}_w}\right) \qquad [20]$$

Note that $\bar{\bar{S}}_w$ will always be greater than or equal to ${}^{\Delta}\bar{S}_w$. If $\bar{\bar{S}}_w$ is equal to ${}^{\Delta}\bar{S}_w$, then the IDC is being followed. For any intermediate point on the MWC, eqn [20] reduces to:

$$\bar{S}_{at} = {}^{I}\bar{S}_{ar}\,{}^{I}\bar{\bar{S}}_w \qquad [21]$$

because ${}^{\Delta}\bar{S}_{\mathrm{w}} = 0$ at the residual water content. Using eqns [18–20], $\bar{S}_{\mathrm{at}}$ can be decoupled from $\bar{\bar{S}}_{\mathrm{w}}$, or ${}^{\mathrm{I}}\bar{\bar{S}}_{\mathrm{w}}$, to calculate θ.

To obtain drying and wetting scanning curves, the IDC (eqn [16]) and MWC (eqn [17]) are scaled through two pertinent reversal points. For drying scanning paths, the IDC is scaled through the saturation path reversal from the previous wetting path to the current drying path and the preceding drying-to-wetting reversal. The equation for predicting drying scanning paths is:

$$\bar{\bar{S}}_{\mathrm{w}}(P_{\mathrm{c}}) = \left\{ \frac{\left[{}^{\mathrm{D}}\bar{\bar{S}}_{\mathrm{w}}(P_{\mathrm{c}}) - {}^{\mathrm{D}}\bar{\bar{S}}_{\mathrm{w}}\left({}^{\mathrm{DI}}P_{\mathrm{c}}\right)\right]\left[{}^{\mathrm{ID}}\bar{\bar{S}}_{\mathrm{w}} - {}^{\mathrm{DI}}\bar{\bar{S}}_{\mathrm{w}}\right]}{\left[{}^{\mathrm{D}}\bar{\bar{S}}_{\mathrm{w}}({}^{\mathrm{ID}}P_{\mathrm{c}}) - {}^{\mathrm{D}}\bar{\bar{S}}_{\mathrm{w}}({}^{\mathrm{DI}}P_{\mathrm{c}})\right]} \right\} + {}^{\mathrm{DI}}\bar{\bar{S}}_{\mathrm{w}} \qquad [22]$$

in which ${}^{\mathrm{D}}\bar{\bar{S}}_{\mathrm{w}}({}^{\mathrm{DI}}P_{\mathrm{c}})$ is the apparent water saturation of the IDC at the capillary pressure of the last drying-to-wetting reversal point (${}^{\mathrm{DI}}P_{\mathrm{c}}$), ${}^{\mathrm{D}}\bar{\bar{S}}_{\mathrm{w}}({}^{\mathrm{ID}}P_{\mathrm{c}})$ is the apparent water saturation of the IDC at the capillary pressure of the last wetting-to-drying reversal point (${}^{\mathrm{ID}}P_{\mathrm{c}}$), and ${}^{\mathrm{DI}}\bar{\bar{S}}_{\mathrm{w}}$ and ${}^{\mathrm{ID}}\bar{\bar{S}}_{\mathrm{w}}$ are the apparent water saturations at previous drying-to-wetting and wetting-to-drying reversal points, respectively.

For wetting scanning paths, the MWC is scaled through the saturation path reversal from the previous drying path to the current wetting path and the preceding wetting-to-drying reversal. The equation for predicting wetting scanning paths is:

$$\bar{\bar{S}}_{\mathrm{w}}(P_{\mathrm{c}}) = \left\{ \frac{\left[{}^{\mathrm{I}}\bar{\bar{S}}_{\mathrm{w}}(P_{\mathrm{c}}) - {}^{\mathrm{I}}\bar{\bar{S}}_{\mathrm{w}}\left({}^{\mathrm{ID}}P_{\mathrm{c}}\right)\right]\left[{}^{\mathrm{DI}}\bar{\bar{S}}_{\mathrm{w}} - {}^{\mathrm{ID}}\bar{\bar{S}}_{\mathrm{w}}\right]}{\left[{}^{\mathrm{I}}\bar{\bar{S}}_{\mathrm{w}}({}^{\mathrm{DI}}P_{\mathrm{c}}) - {}^{\mathrm{I}}\bar{\bar{S}}_{\mathrm{w}}({}^{\mathrm{ID}}P_{\mathrm{c}})\right]} \right\} + {}^{\mathrm{ID}}\bar{\bar{S}}_{\mathrm{w}} \qquad [23]$$

in which ${}^{\mathrm{I}}\bar{\bar{S}}_{\mathrm{w}}({}^{\mathrm{ID}}P_{\mathrm{c}})$ is the apparent water saturation of the MWC at the capillary pressure of the last wetting-to-drying reversal point, and ${}^{\mathrm{I}}\bar{\bar{S}}_{\mathrm{w}}({}^{\mathrm{DI}}P_{\mathrm{c}})$ is the apparent water saturation of the MWC at the capillary pressure of the last drying-to-wetting reversal point. For wetting paths from the IDC, the preceding drying-to-wetting reversal apparent water saturation equals 1.

The scaling assumes that the scanning paths have the same general curvature as either the IDC or the MWC. Furthermore, the scaling enforces that the scanning curves close at the reversal points. Using the model of Lenhard *et al.*, an infinite number of saturation paths can be modeled from the IDC and MWC parameters and the reversal-point saturations. Using eqns [18–20], $\bar{S}_{\mathrm{at}}$ can be decoupled from the $\bar{\bar{S}}_{\mathrm{w}}$ calculated with either eqn [22] or [23] to yield θ.

To account for effects of entrapped air on the conduction of water, Lenhard and Parker have modified an equation frequently used by soil scientists to predict the ability of the pore spaces to conduct water. As previously discussed for water retention, they utilize the concept of apparent saturation and make a correction for the pore space containing entrapped air. Their equation to predict water relative permeability, k_{rw}, which is a measure of the ability of the pore space to conduct water relative to when all of the pore spaces are filled with water, is:

$$k_{\mathrm{rw}} = \bar{S}_{\mathrm{w}}^{0.5} \left[\frac{\displaystyle\int_{0}^{\bar{\bar{S}}_{\mathrm{w}}} \frac{\mathrm{d}\bar{S}_{\mathrm{e}}}{P_{\mathrm{c}}\left(\bar{S}_{\mathrm{e}}\right)} - \int_{0}^{\bar{S}_{\mathrm{at}}} \frac{\mathrm{d}\bar{S}_{\mathrm{at}}}{P_{\mathrm{c}}\left(\bar{S}_{\mathrm{e}}\right)}}{\displaystyle\int_{0}^{\bar{\bar{S}}_{\mathrm{w}}} \frac{\mathrm{d}\bar{S}_{\mathrm{e}}}{P_{\mathrm{c}}\left(\bar{S}_{\mathrm{e}}\right)}} \right] \qquad [24]$$

where, for an air–water system, $\bar{S}_{\mathrm{e}} = \bar{\bar{S}}_{\mathrm{w}}$ is the proportion of the effective porosity in which water flow can potentially occur; and $P_{\mathrm{c}}(\bar{\bar{S}}_{\mathrm{e}})$ is the inverse of the function describing the IDC (eqn [16]). The effect of eqn [24] is to yield a larger k_{rw} for scanning curves than for the IDC at the same water content, because the entrapment of air forces water to move into the larger pores, allowing for a greater potential to conduct water. Typically, soil scientists have ignored hysteresis in relations between k_{rw} and water content because the differences are relatively minor at low-to-moderate water-content values. However, the effects of air entrapment, one of the hysteresis components, may yield significant differences in k_{rw} relations at high-water-content values.

It is important to consider the effects of hysteresis on water behavior in the subsurface. A fluctuating water table experiment has been conducted in which water pressures and saturations were measured. A computer model that either accounts for or ignores hysteresis in the saturation–capillary pressure–relative hydraulic conductivity relations is used to simulate the experiment. Comparisons of simulation and experimental results are shown in **Figures 3 and 4**. In **Figure 3**, water saturation values are plotted as a function of time for an elevation of 60 cm above the lower boundary of the experimental domain, which has a height of 72 cm. The experiment started with the water table at the surface. During a 2-h period, the water table was decreased 5 cm every 10 min to an elevation of 7 cm above the lower boundary of the experimental domain. Starting at $t = 3$ h, the water table was subsequently increased 5 cm every 10 min until an elevation of 42 cm had been reached. This increase was then followed again by a decrease at $t = 5$ h in the water table elevation to 17 cm at the same rate as before. Finally, at $t = 6.67$ h, the water table was

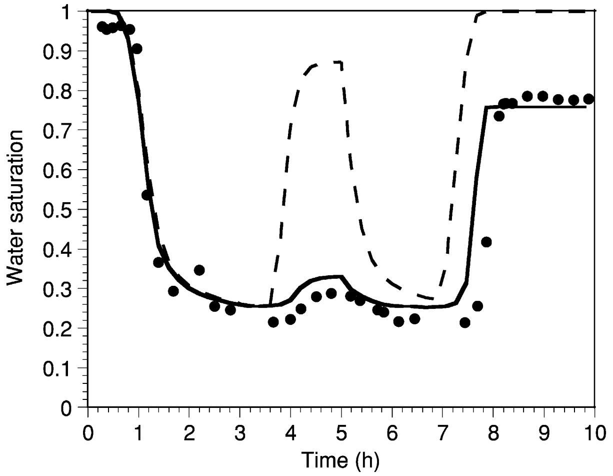


Figure 3 Comparison of measured (solid circles) with simulated fully hysteretic (solid line) and simulated nonhysteretic water content values during water table fluctuations at 60 cm above the bottom boundary of the flow domain.

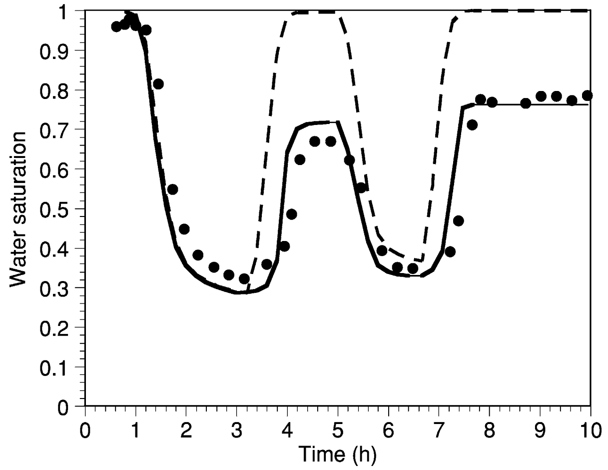


Figure 4 Comparison of measured (solid circles) with simulated fully hysteretic (solid line) and simulated nonhysteretic water content values during water table fluctuations at 70 cm above the bottom boundary of the flow domain.

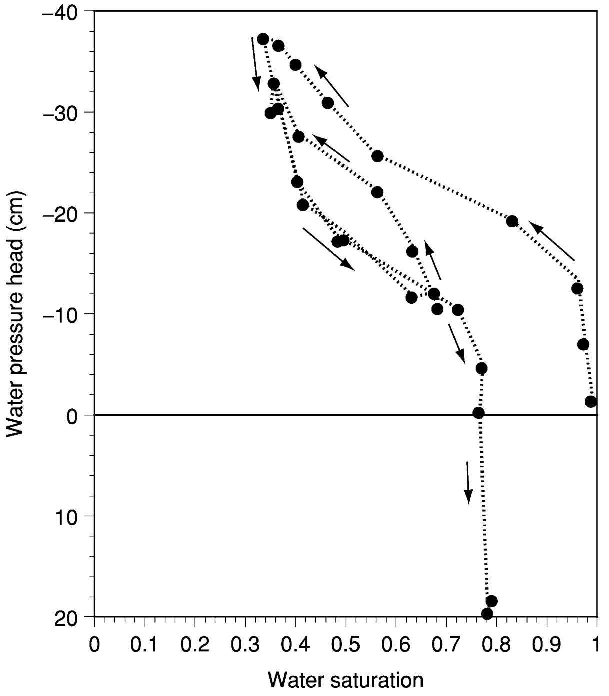


Figure 5 Measured saturation–matric-head (in this case equal to the pressure head) history at 50 cm above the bottom boundary of the flow domain showing a closed internal scanning loop. The solid circles represent the data points, while the arrows indicate the direction in which the water content is changing.

increased to the surface at the same rate as before. It can be seen that increasing the water table from an elevation of 7 to 42 cm, and subsequently lowering it from 42 to 17 cm had only a small effect on the measured water content values ($3.5 < t < 6$ h). It was not until later, when the water table was raised from 17 cm to the soil surface, that a significant increase in water content occurred. For $8.5 < t < 10$ h, the water table is located above the 60-cm measurement elevation and the water content remains constant with time, as is expected for a strongly water-wet soil. The water saturation does not approach 1 during this time, because air was entrapped in the pore spaces, which constitute slightly more that 20% of the pore volume. Note that, over the short duration of the experiment, the entrapped air does not move or dissolve into the water. In Figure 4, the results are shown for an elevation of 50 cm above the lower boundary. Different water content values from those in Figure 3 were measured because the saturation path history is different than for the data in Figure 3. In fact, all elevations will possess a different saturation path history. The results shown in Figures 3 and 4 demonstrate that the simulation results obtained using an algorithm that accounts for hysteresis are significantly better than those when hysteresis is ignored. Figure 5 shows the measurements of the soil water matric-head versus saturation at the 50-cm elevation. The closed scanning loop was produced from the first raising of the water-table elevation to 42 cm and the subsequent lowering of the water table to 17 cm.

Summary

Hysteresis in water-retention curves causes these curves to be different during wetting than during drying, i.e., at the same water-content values, different values for the soil water matric head or capillary pressure exist or vice versa. Ignoring hysteresis when predicting soil water content profiles by means of simulation models results in erroneous information.

See also: **Capillarity**; **Drainage, Surface and Subsurface**; **Infiltration**; **Leaching Processes**; **Porosity and Pore-Size Distribution**; **Vadose Zone:** Hydrologic Processes; **Water Potential**

Further Reading

Bomba SJ (1968) *Hysteresis and Time-Scale Invariance in a Glass-Bead Medium*. PhD Thesis. Madison, WI: University of Wisconsin Press.

Dane JH and Klute A (1977) Salt effects on the hydraulic properties of a swelling soil. *Soil Science Society of America Journal* 41: 1043–1049.

Dane JH and Wierenga PJ (1975) Effect of hysteresis on the prediction of infiltration, redistribution and drainage of water in a layered soil. *Journal of Hydrology* 25: 229–242.

Haines WB (1930) Studies in the physical properties of soils. V. The hysteresis effect in capillary properties and the modes of moisture distribution associated therewith. *Journal of Agricultural Science* 20: 97–116.

Hillel D (1980) *Fundamentals of Soil Physics*. New York: Academic Press.

Hillel D and Mottes J (1966) Effect of plate impedance, wetting method and aging on soil moisture retention. *Soil Science* 102: 135–140.

Koorevaar P, Menelik G, and Dirksen C (1983) *Element of Soil Physics*. Developments in soil science 13. Elsevier Science.

Land CS (1968) Calculation of imbibition relative permeability for two- and three-phase flow from rock properties. *Society of Petroleum Engineering Journal* 8: 149–156.

Lenhard RJ (1992) Measurement and modeling of three-phase saturation-pressure hysteresis. *Journal of Contaminant Hydrology* 9: 243–269.

Lenhard RJ and Parker JC (1987) A model for hysteretic constitutive relations governing multiphase flow. 2. Permeability–saturation relations. *Water Resources Research* 23: 2197–2206.

Lenhard RJ, Parker JC, and Kaluarachchi JJ (1991) Comparing simulated and experimental hysteretic two-phase transient fluid flow phenomena. *Water Resources Research* 27: 2113–2124.

Miller EE and Miller RD (1955a) Theory of capillary flow. I. Practical implications. *Soil Science Society of America Proceedings* 19: 267–271.

Miller EE and Miller RD (1955b) Theory of capillary flow. II. Experimental information. *Soil Science Society of America Proceedings* 19: 271–275.

Miller EE and Miller RD (1956) Physical theory for capillary flow phenomena. *Journal of Applied Physics* 27: 324–332.

Parker JC and Lenhard RJ (1987) A model for hysteretic constitutive relations governing multiphase flow. 1. Saturation-pressure relations. *Water Resources Research* 23: 2187–2196.

Philip JR (1964) Similarity hypothesis for capillary hysteresis in porous materials. *Journal of Geophysics Research* 69: 1553–1562.

Topp GC (1969) Soil water hysteresis measured in a sandy loam and compared with the hysteresis domain model. *Soil Science Society of America Proceedings* 33: 645–651.

Topp GC and Miller EE (1966) Hysteresis moisture characteristics and hydraulic conductivities for glass-bead media. *Soil Science Society of America Proceedings* 30: 156–162.

van Genuchten MTh (1980) A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Science Society of America Journal* 44: 892–898.

I

IMMISCIBLE FLUIDS

R J Lenhard, Idaho National Engineering Environmental Laboratory, Idaho Falls, ID, USA
J H Dane, Auburn University, Auburn, AL, USA
M Oostrom, Pacific Northwest National Laboratory, Richland, WA, USA

Published by Elsevier Ltd.

Introduction

Soils are located near the surfaces of land on Earth. As such, they contain variable amounts of air and water, which are immiscible fluids. Soils are also subject to contamination by human activities. Many of the contaminants commonly found in soils are liquids that are immiscible with air and water. These immiscible contaminants are principally fuel hydrocarbons and industrial solvents, and are referred to as nonaqueous-phase liquids (NAPLs). Fuel hydrocarbons (petroleum liquids) tend to be NAPLs that are less dense than water (LNAPLs). Industrial solvents, especially chlorinated hydrocarbons, tend to be denser than water (DNAPLs). The increased interest in better understanding NAPL behavior in soils is related to our concern for water quality. As water moves downward through soils, dissolved NAPL components can migrate to potable water supplies. Furthermore, high vapor-pressure components of NAPLs may partition to the soil air and move through the soil, thereby mixing the volatilized components with the soil air. These volatilized components can later partition to potable water supplies. The migration rate of the dissolved and volatile NAPL components and the NAPL movement toward groundwater are affected by the distribution of NAPL in soils.

The distribution and movement of air, water, and NAPLs in soils are governed by differences in fluid pressures among the fluid phases, the gradient of pressure in each fluid phase, and the size of the pores containing the fluids. The partitioning of components among the fluid phases may vary spatially and temporally. Overall, the behavior of NAPLs in soils is a consequence of complex physical, chemical, and biological processes. In the following sections, we outline some of the processes.

Fluid Distribution in the Pore Spaces

The distribution of air, water, and NAPL in pore spaces is a function of their pressures and the molecular forces of attraction for each other and for the surfaces of the soil particles. ‘Wettability’ describes the propensity of fluids for solid surfaces and other fluids. It is a complex function of fluid and solid properties. One measure of wettability is the contact angle that a fluid makes with a solid surface in the presence of other fluids. The wettability will determine whether fluids will occupy larger or smaller pores in soils. A ‘wetting fluid’ means that the fluid will spread over solid surfaces and a ‘nonwetting fluid’ means that the fluid is not in contact with solids. In soils, water is generally accepted to be the wetting fluid, and NAPL and air are nonwetting fluids. It is also generally accepted that NAPL will ‘wet’ (spread over) water surfaces. Therefore, water typically occupies the smallest pore spaces in soils and air occupies the largest pore spaces. NAPL occupies pore spaces intermediate to those occupied by water and air. However, there are several exceptions to these conditions. One exception occurs when organic compounds coat soil particles. Another exception is when the NAPL has a complex structure in which one end of the compound has a hydrophilic characteristic and the opposite end has a hydrophobic characteristic. For the former exception, the NAPL will probably wet the organic-coated soil particles instead of water. For the latter exception, a condition of mixed wettability may develop that can result in water occupying pores smaller and larger than those occupied by NAPL during water imbibition.

A measure of the molecular forces of attraction between immiscible fluids is the interfacial tension, which is a macroscopic and statistical property of the interface separating two immiscible fluids. The interfacial tension is a measure of the free energy per unit area of the interface. The SI units for interfacial tension are force per unit length (newtons per meter). When one of the fluids is a gas and the other a liquid, then the term 'surface tension' is commonly used instead of 'interfacial tension.' Surface tension and interfacial tension are different words for the same phenomenon. In soils, the distribution of immiscible fluids is different for identical pore geometries and fluid pressures if the interfacial tensions are different.

A basic equation of capillarity is the Young–Laplace equation, which was proposed in 1805. The equation relates the difference in fluid pressures across the interface separating immiscible fluids to the interfacial tension and the radii of curvature of the interface at mechanical equilibrium. The Young–Laplace equation is:

$$\Delta P = \sigma\left(\frac{1}{R_1} + \frac{1}{R_2}\right) \qquad [1]$$

in which ΔP (pascals) is the difference in pressures of the fluids across the interface, σ (newtons per meter) is the interfacial tension of the fluid pair, and R_1 and R_2 (meters) are radii of curvature of the interface in orthogonal directions. In soils, ΔP is commonly called the capillary pressure, which is defined by:

$$P_c = P_{nw} - P_w \qquad [2]$$

in which P_c (pascals) is the capillary pressure, P_{nw} (pascals) is the nonwetting-fluid pressure, and P_w (pascals) is the wetting-fluid pressure. The interface separating the nonwetting and wetting fluids is curved convex toward the nonwetting fluid, which means that the pressure of the nonwetting fluid is greater than the wetting fluid. Consequently, the capillary pressure is always positive using the definition given by eqn [2]. The radii of curvature of the fluid–fluid interface in eqn [1] can be converted to radii of soil pores by:

$$\cos\alpha = \frac{r}{R} \qquad [3]$$

in which α is the contact angle, r is a radius of the soil pore, and R is a radius of the fluid–fluid interface curvature.

In air–NAPL–water systems in soils, there are two capillary pressures. When water is the wetting fluid and NAPL wets water surfaces, there are interfaces separating the water from the NAPL and the NAPL from the air. The pressure difference across the NAPL–water interfaces is called the NAPL–water capillary pressure. The pressure difference across the air–NAPL interfaces is called the air–NAPL capillary pressure. For a given pore geometry in soils, there is a relationship between capillary pressures and fluid contents. One measure of the fluid content is the ratio of the volume of fluid (cubic meters) per unit volume of soil (cubic meters), i.e.:

$$\theta_i = \frac{V_i}{V_t} \qquad [4]$$

in which θ_i is the volumetric fluid content of fluid i and V_i is the volume of fluid i in the total volume of soil (V_t), including solid and void spaces. Another measure of fluid content is the ratio of the volume of fluid (cubic meters) per unit of void volume in a soil sample (cubic meters) i.e.:

$$S_i = \frac{V_i}{V_v} \qquad [5]$$

in which S_i is the saturation of fluid i and V_v is the total void volume in the soil sample containing the volume of fluid i. The volumetric fluid content and the fluid saturation are related to the porosity of the soil through:

$$\phi = \frac{\theta_i}{S_i} = \frac{V_v}{V_t} \qquad [6]$$

in which ϕ is the porosity of the soil. Because the measures of fluid content and porosity are expressed as ratios of volumes, they are dimensionless.

The air–water capillary pressure (P_{aw}) and the water, or aqueous-phase, saturation (S_{aq}) are related to each other in an air–water system. The NAPL–water capillary pressure (P_{ow}) and the aqueous-phase saturation are related to each other in an air–NAPL–water system. The air–NAPL capillary pressure (P_{ao}) and total-liquid saturation (S_l) are related to each other in an air–NAPL–water system. The total-liquid saturation is the sum of the NAPL and aqueous-phase saturations.

There have been several proposed models to describe mathematically the relations between capillary pressures and fluid saturations. The two most common models for describing *S*–*P* relations are those by Brooks and Corey, and van Genuchten. The models were initially proposed for describing *S*–*P* relations in air–water systems. Since then they have been adapted to describe *S*–*P* relations in air–NAPL–water systems. The Brooks–Corey model for an air–water system is:

$$\begin{aligned} \bar{S}_{aq} &= 1 \quad \text{for } P_{aw} \le P_d^{aw} \\ \bar{S}_{aq} &= \left(\frac{P_d^{aw}}{P_{aw}}\right)^{\lambda} \quad \text{for } P_{aw} > P_d^{aw} \end{aligned} \qquad [7]$$

where

$$\bar{S}_{aq} = \frac{S_{aq} - S_r}{1 - S_r}$$

in which $\bar{S}_{aq}$ is the effective aqueous-phase saturation, P_d^{aw} is the displacement pressure in an air–water system, λ is a parameter related to the pore-size distribution, and S_r is the residual aqueous-phase saturation.

The van Genuchten model for an air–water system is:

$$\bar{S}_{aq} = [1 + (\alpha_{aw} P_{aw})^n]^{-m} \quad \text{for } P_{aw} \geq 0 \qquad [8]$$

where

$$m = \frac{n-1}{n}$$

and α_{aw} and n are curve-fitting parameters.

The extension of the above equations to describe *S–P* relations in air–NAPL–water systems was proposed by Parker *et al.* To minimize the number of parameters needed to describe three-fluid *S–P* relations, a scaling parameter was used. A scaled *S–P* function was defined that would apply to air–water and air–NAPL–water systems. An implicit assumption in the approach is that the effects of clay–fluid interactions on *S–P* relations are similar in air–NAPL–water systems to those in air–water systems. The scaling format used by Parker *et al.*, which has similarities to scaling methods proposed in the petroleum industry, is:

$$\begin{aligned}
\bar{S}^{aw}_{aq}(\beta_{aw} P_{aw}) &= S^*(P^*) \\
\bar{S}^{ow}_{aq}(\beta_{ow} P_{ow}) &= S^*(P^*) \\
\bar{S}^{ao}_{o}(\beta_{ao} P_{ao}) &= S^*(P^*) \\
\bar{S}^{aow}_{aq}(\beta_{ow} P_{ow}) &= S^*(P^*) \\
\bar{S}^{aow}_{l}(\beta_{ao} P_{ao}) &= S^*(P^*)
\end{aligned} \qquad [9]$$

in which $\bar{S}^{aw}_{aq}$ and $\bar{S}^{aow}_{aq}$ are the effective aqueous-phase saturations in the air–water and in the air–NAPL–water fluid systems, respectively; $\bar{S}^{aow}_{l}$ is the effective total-liquid saturation in the air–NAPL–water system; $\bar{S}^{ow}_{aq}$ is the effective aqueous-phase saturation in a two-phase NAPL–water system; $\bar{S}^{ao}_{o}$ is the effective NAPL saturation in a two-phase air–NAPL system; β_{aw}, β_{ow}, and β_{ao} are the air–water, NAPL–water, and air–NAPL scaling factors, respectively; P_{aw}, P_{ow}, and P_{ao} are the air–water, NAPL–water, and air–NAPL capillary pressures, respectively; and $S^*(P^*)$ is the scaled *S–P* relation. The scaling format given in eqn [9] has been tested for nonhysteretic *S–P* data (i.e., main drainage branch *S–P* relations) and the results (**Figure 1**) validated the basic concepts of the scaling format.

Describing NAPL movement in soils requires that air–water and air–NAPL–water systems be considered simultaneously in any model. As NAPL moves into a soil or as it drains, volatilizes, or dissolves, there must be no discontinuities in fluid contents as a function of capillary pressure. To prevent any such discontinuities, the scaling factors must satisfy:

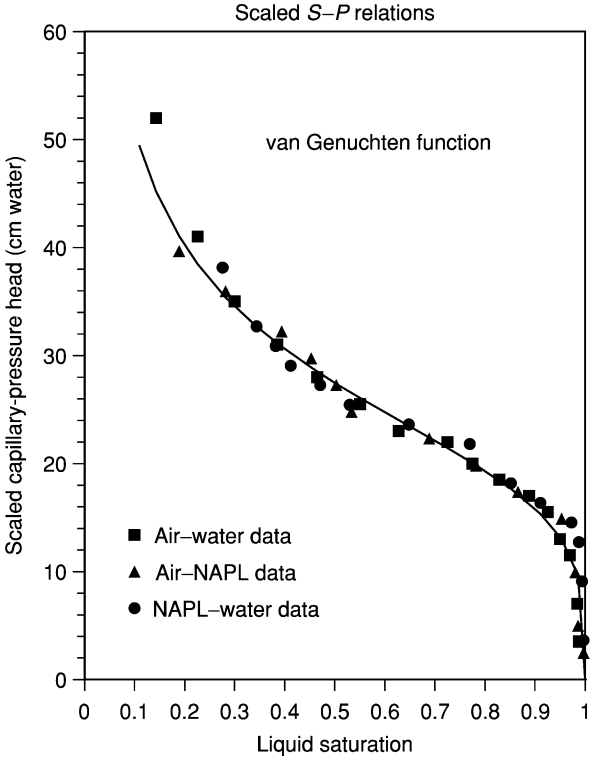


Figure 1 Scaling of capillary pressure–saturation measurements of different fluid systems using Lenhard's scaling format. (Adapted from Lenhard RJ (1994) Scaling fluid content–pressure relations of different fluid systems in porous media. In: Morel-Seytoux HJ (ed.) *Proceedings of 14th Hydrology Days*, pp. 223–235. Atherton, CA: Hydrology Days Publications, with permission.)

$$\frac{1}{\beta_{aw}} = \frac{1}{\beta_{ow}} + \frac{1}{\beta_{ao}} \qquad [10]$$

The scaling factors can be determined from *S–P* data in two-fluid or three-fluid systems or predicted from ratios of interfacial tensions. The β_{ow} and β_{ao} scaling factors are determined when the $\bar{S}^{aow}_{aq}(\beta_{ow} P_{ow})$ and $\bar{S}^{aow}_{l}(\beta_{ao} P_{ao})$ or $\bar{S}^{ow}_{aq}(\beta_{ow} P_{ow})$ and $\bar{S}^{ao}_{aq}(\beta_{ao} P_{ao})$ relations, respectively, are scaled to the $\bar{S}^{aw}_{aq}(\beta_{aw} P_{aw})$ relations in which β_{aw} is defined as 1. To ensure that eqn [10] is followed when β_{ow} and β_{ao} are determined from regression analyses of *S–P* data, β_{ao} should be set to:

$$\beta_{ao} = \frac{\beta_{ow}}{\beta_{ow} - 1} \qquad [11]$$

By following this procedure, then only one scaling factor (β_{ow}) needs to be determined in the regression analyses. If $\bar{S}^{aow}_{aq}(\beta_{ow} P_{ow})$ and $\bar{S}^{aow}_{l}(\beta_{ao} P_{ao})$ relations are unavailable, then the β_{ao} and β_{ow} scaling factors can be estimated from ratios of air–NAPL and NAPL–water interfacial tensions. To ensure that no discontinuities in fluid contents exist as a function of capillary pressure, the β_{ao} and β_{ow} scaling factors can be determined from:

$$\beta_{\text{ao}} = \frac{\sigma_{\text{ao}} + \sigma_{\text{ow}}}{\sigma_{\text{ao}}}$$
$$\beta_{\text{ow}} = \frac{\sigma_{\text{ao}} + \sigma_{\text{ow}}}{\sigma_{\text{ow}}} \qquad [12]$$

Figure 2 shows the distribution of air, NAPL, and water above the water-saturated region in a soil when the scaling factors conform to eqn [10] and when they do not. The discontinuity in water content at the transition between the air–water and the air–NAPL–water regions can be seen in **Figure 2b** when eqn [10] is not followed.

The relationship between fluid content and capillary pressure is very complex and not unique; it depends on the history of fluid movements. This phenomenon is called hysteresis. In air–water systems, there are only two possible combinations of fluid movement. Either water is imbibing and air is draining, or water is draining and air is imbibing. In air–NAPL–water systems, there are six possible combinations of fluid movement, which makes predicting, or modeling, three-fluid *S–P* behavior more difficult. Because of the difficulty of describing hysteresis, many modelers neglect it.

Major reasons for hysteresis in *S–P* relations are fluid entrapment, pore-geometry effects, and contact angle changes with direction of fluid movement (draining versus imbibing). As nonwetting fluid is displaced by more wetting fluid, the nonwetting fluid can become trapped and isolated in the soil pores by various mechanisms. The net result causes the wetting fluid to occupy larger pores on imbibition saturation paths than it would on drainage saturation paths for the same wetting-fluid saturation. When the fluid–fluid interfaces are in larger pores, the radii of curvature of the interfaces will be larger (see eqn [3]), which means the capillary pressure will be smaller (see eqn [1]). Pore-geometry effects include the phenomenon that some pores initially filled with fluid may not refill at the same capillary pressure at which they drained because the direction of fluid movement is different. Sometimes this phenomenon is called the ink-bottle effect. Contact angle effects result because the angle a fluid makes with the soil particle surfaces may be different depending if the fluid is advancing or retreating over the soil or fluid surfaces. A different contact angle for the same fluid content will yield a different capillary pressure. However, the change in contact angles in soils with direction of fluid movement is very complex, because the soil particle surfaces are very irregular at the microscopic scale. The magnitude of hysteresis can vary significantly in soils.

Parker and Lenhard have proposed a method to model hysteretic air–NAPL–water *S–P* relations in which water is the wetting fluid. An apparent

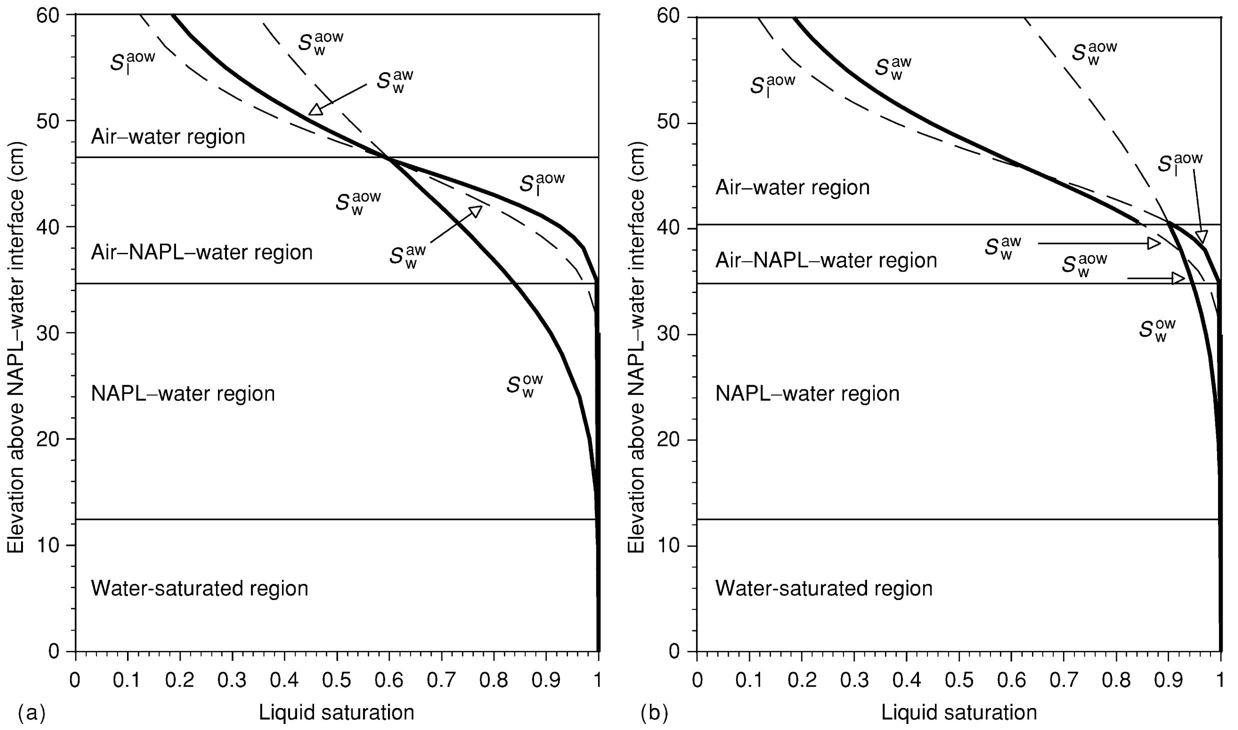


Figure 2 Distributions of air nonaqueous-phase liquids (NAPL), and water above the water-saturated region in which (a) the predictions were obtained using eqns [10] and [12] and (b) the predictions were obtained by only best-fitting the pertinent *S–P* data. Reproduced with permission from Lenhard RJ, Oostrom M, and Dane JH (2002) Prediction of capillary pressure-relative permeability relations. In: Dane JH and Topp GC (eds) Methods of Soils Analysis, Part 4 – Physical Methods. Soil Science Society of America Book Series 5.

saturation was defined to account for fluid entrapment. For an air–water system, the entrapped air in water can be determined from the apparent aqueous-phase saturation and the aqueous-phase saturation-path history. For an air–NAPL–water system, the entrapped NAPL in water can be determined from the apparent aqueous-phase saturation and the aqueous-phase saturation-path history. Also, for an air–NAPL–water system, the entrapped air in NAPL and water can be determined from the apparent total-liquid and aqueous-phase saturations and the saturation-path histories.

Parker and Lenhard have defined and parameterized two main *S*–*P* paths that describe relations between apparent saturations and capillary pressures. One path is for main drainage *S*–*P* relations, and the other path is for main imbibition relations. To describe *S*–*P* relations between the NAPL–water capillary pressure and the apparent aqueous-phase saturation, and relations between the air–NAPL capillary pressure and the apparent total-liquid saturation, the two main *S*–*P* paths of an air–water system and the scaling format given in eqn [9] are used. To describe any drainage *S*–*P* path, the main drainage *S*–*P* relations are scaled to pass through the pertinent saturation-path reversal points (imbibition to drainage and drainage to imbibition). To describe any imbibition *S*–*P* path, the main imbibition *S*–*P* relations are scaled to pass through the pertinent saturation-path reversal points. This approach has been modified to allow concurrent entrapment of NAPL and air by water. The hysteretic *S*–*P* model has also been tested against experimental data. In Figure 3, the air-water and air–NAPL–water *S*–*P* model is tested against the water saturation-path experimental data. In Figure 4, the *S*–*P* model is tested against the total-liquid saturation-path experimental data. In both cases, the model is able to describe the experimental data very well.

Recent modifications to the hysteretic *S*–*P* model involve accounting for residual NAPL in an air–NAPL–water system. Experimental investigations have shown that some NAPL will not drain from the soil pores in the vadose zone, even after long periods of time. The NAPL that does not drain from the soil pores is assumed to be immobile and has been called residual NAPL. Modifications to the hysteretic air–NAPL–water *S*–*P* model have been made to account for residual NAPL. The total NAPL saturation was defined to consist of mobile (free), immobile (residual), and water-trapped components. The amount of each component is calculated from the capillary pressures and the saturation-path histories.

Movement of Immiscible Fluids in Soils

The dimensions of the soil pores containing fluids are a very important factor for determining how readily those fluids move throughout the pore spaces in a

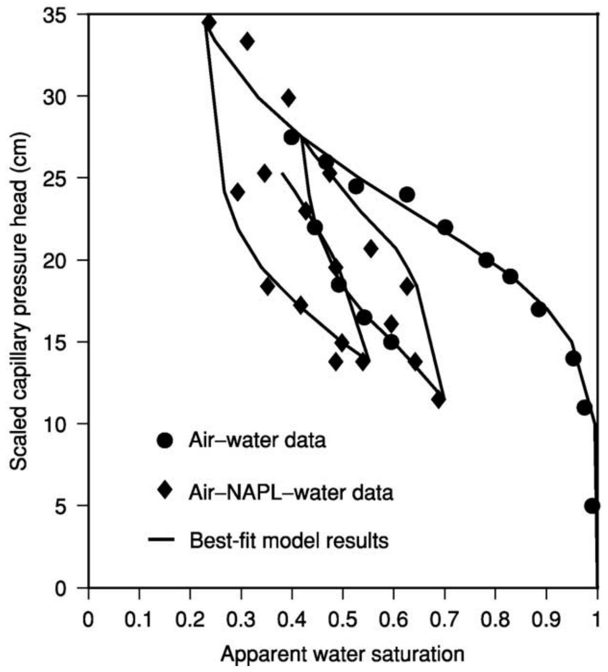


Figure 3 Scaling hysteretic water *S*–*P* relations of different fluid systems. (Adapted from Lenhard RJ (1992) Measurement and modeling of three-phase saturation–pressure hysteresis. *Journal of Contaminant Hydrology* 9: 243–269.)

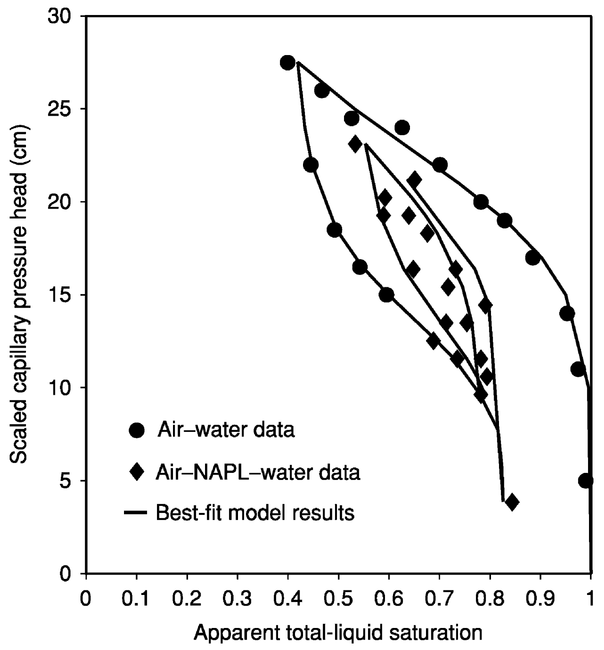


Figure 4 Scaling hysteretic total-liquid *S*–*P* relations of different fluid systems. (Adapted from Lenhard RJ (1992) Measurement and modeling of three-phase saturation–pressure hysteresis. *Journal of Contaminant Hydrology* 9: 243–269.)

soil. The Poiseuille equation shows that the velocity of fluid flowing through a circular tube varies with the square of the radius of the tube. Therefore, the fluid velocity in a tube having twice the radius of another tube is 4 times as great for the same gradient in piezometric pressure. The latter is an important factor governing how readily fluids move throughout the pore spaces in a soil. Because the geometry of soil pores is far more complicated than circular tubes, calculating the fluid velocity is more complex than the Poiseuille equation. However, the same principles apply in that larger soil pores are able to conduct fluid with less resistance than smaller pores.

The description of multifluid flow in soils is based on a modification of the Darcy law for saturated water flow, which was developed in 1856. In 1907, Buckingham modified the Darcy law to describe unsaturated water flow by replacing the saturated hydraulic conductivity with an unsaturated hydraulic conductivity that is a function of water content or capillary pressure. Over the years, the modification made by Buckingham and the recognition that the hydraulic conductivity consists of fluid-dependent and -independent elements has yielded what is commonly referred to as a generalized form of the Darcy law:

$$q_\mathrm{i} = -\frac{k_\mathrm{ri}k}{\mu_\mathrm{i}}\left(\frac{\partial P_\mathrm{i}}{\partial x_\mathrm{j}} + \rho_\mathrm{i}\mathbf{g}\right) \qquad [13]$$

in which q_i (meters per second) is the volumetric flux density of fluid i per cross-sectional area of a porous medium, k_ri is the relative permeability of fluid i, k (square meters) is the intrinsic permeability, μ_i (newton-seconds per square meter) is the viscosity of fluid i, P_i is the pressure of fluid i, ρ_i (kilograms per cubic meter) is the mass density of fluid i, $\mathbf{g}$ (meters per second squared) is the gravitational vector, and x_j (meters) is a symbol representing the Cartesian coordinates x, y, and z. The term in brackets in eqn [13] is the gradient of piezometric pressure. The relative permeability given in eqn [13] is a ratio of the permeability (square meters) at the relevant fluid saturation (i.e., effective permeability) divided by the intrinsic permeability.

A key parameter in eqn [13] that is very difficult and time-consuming to measure is the relative permeability, k_r. Almost all relative permeability measurements in air–NAPL–water systems have been conducted when the water saturation is at its residual value and for specific saturation paths. In the petroleum industry, it is common to measure the NAPL (oil) and water relative permeabilities for saturation paths similar to those that will be used during oil recovery. The measurements are made on samples that are carefully extracted from the petroleum reservoir and stored so as not to change the wettability of the samples. Measurements are generally conducted at temperatures and pressures of the petroleum reservoir. In determining cleanup strategies for NAPL-contaminated soils, measurements of relative permeability are typically not conducted, because they are expensive and no standard measurement techniques have been developed for soils. Whereas the water content in petroleum reservoirs is commonly at the residual water saturation, the water content in soils is usually much higher. Therefore, air–NAPL–water relative permeability measurement techniques used in the petroleum industry are not applicable to soils. Additionally, the cost for the elaborate relative permeability measurements in the petroleum industry cannot be justified for conducting analyses of NAPL-contaminated soils. Relative permeabilities that are needed to predict fluid flow of air–NAPL–water systems in soils are normally predicted.

The stationary fluid method has been applied to determine NAPL effective permeability values in NAPL–water systems. The method allows for direct measurements during steady-state flow conditions (constant capillary pressure and fluid saturations) throughout the sample. After each steady-state flow condition, the water saturation can be either increased or decreased.

Prediction of relative permeabilities is based on theory developed by petroleum engineers and soil scientists. In 1949, Purcell formulated a method to predict relative permeability using the porosity and *S*–*P* relations. Quickly thereafter, several investigators in the early 1950s developed models for predicting relative permeability using concepts proposed by Kozeny-Carman and Purcell. These early efforts focused on predicting relative permeabilities in two-fluid systems. It was not until the mid-1950s that investigators began to develop models for air–NAPL–water systems.

Investigations in the petroleum industry have shown that the water relative permeability in an air–NAPL–water system depends only on the water saturation. Similar investigations have also shown that the air relative permeability depends only on the air saturation. In an air–NAPL–water system, the air saturation is equal to $1 - S_\mathrm{l}$; therefore, the air relative permeability is also a function of the total-liquid saturation. The NAPL relative permeability has been found to be a complex function of the water and NAPL (or total-liquid) saturations. To demonstrate how the NAPL relative permeability varies with water and NAPL saturations, predictions of the NAPL relative permeability using equations developed by Lenhard and Parker are shown in **Figure 5** for water saturations of 0.2 and 0.5. Hysteresis exists in *k*–*S* relations because of

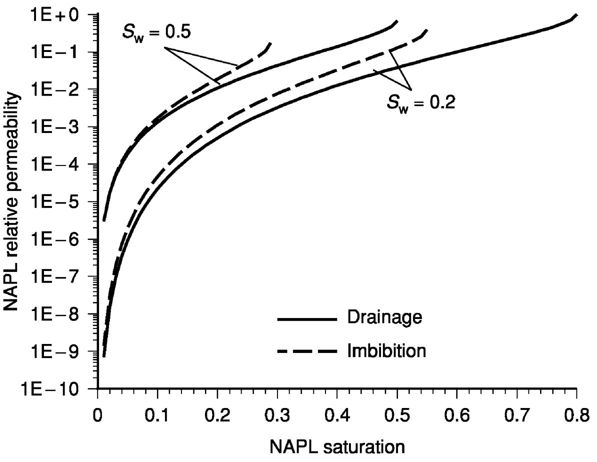


Figure 5 Predicted hysteretic NAPL relative permeability at two different water saturations. (Adapted from Lenhard RJ and Parker JC (1987) A model for hysteretic constitutive relations governing multiphase flow. 2. Permeability–saturation relations. *Water Resources Research* 23: 2197–2206.)

entrapment of fluids with lesser wettability by fluids with greater wettability. For example, air and NAPL can be entrapped and occluded by water, but air can only be entrapped and occluded by NAPL in strongly water-wet porous media. Water theoretically cannot be entrapped by NAPL or air in strongly water-wet porous media, because it is the wetting fluid (i.e., the fluid with the greatest wettability). It can be seen that significantly different NAPL relative permeability relations occur depending on the water and NAPL saturations. Furthermore, there can be large differences in the predicted NAPL relative permeability because of hysteresis. The difference is greater for larger total-liquid saturations.

Partitioning of NAPL Components Among the Fluid and Solid Phases

NAPL components in soils are present in various forms. They can exist in the aqueous, vapor, and solid phases as well as in a distinct liquid phase (i.e., NAPL). To model the subsurface behavior of NAPL components, the partitioning of the components among the liquid, gas, and solid phases needs to be considered. Mass transfer of NAPL components in strongly water-wet porous media involves three main processes: volatilization into the gas phase from the aqueous phase and NAPL; dissolution into the aqueous phase from the gas phase and NAPL; and sorption from the aqueous phase to the solid phase. Because water is always the wetting fluid in strongly water-wet porous media, it coats the soil grains, and gas-to-solid and NAPL-to-solid mass transfer are not possible. NAPL components in the gas phase and the NAPL must first partition to the aqueous phase before sorbing on solid surfaces.

Researchers have applied several different models to describe mass transfer of compounds between aqueous, gas, and solid phases. Because many mass-transfer processes are assumed to occur much faster than the convective flow of the fluids, mass transfer is commonly modeled as being instantaneous and local equilibrium is assumed. The simplest equilibrium model involves a sorption isotherm in which the concentration of a compound in the aqueous phase is linearly related to the concentration adsorbed to the solid surfaces. The so-called linear equilibrium isotherm is described by the equation:

$$C_{\mathrm{s}}^{\mathrm{i}} = K_{\mathrm{d}}^{\mathrm{i}} C_{\mathrm{aq}}^{\mathrm{i}} \qquad [14]$$

in which $C_{\mathrm{s}}^{\mathrm{i}}$ is the mass of the sorbed NAPL component i per unit dry mass of solids (kilograms per kilogram), $K_{\mathrm{d}}^{\mathrm{i}}$ is the partitioning coefficient (cubic meters per kilogram) of NAPL component i, and $C_{\mathrm{aq}}^{\mathrm{i}}$ is the mass of dissolved NAPL component i per unit volume of the aqueous phase (kilograms per cubic meter).

If the organic carbon fraction of soils is greater than 0.01 (kilograms per kilogram), then sorption of dissolved NAPL components is dominated by the organic fraction. In this case, the partitioning coefficient, $K_{\mathrm{d}}^{\mathrm{i}}$, can be estimated by:

$$K_{\mathrm{d}}^{\mathrm{i}} = f_{\mathrm{oc}} K_{\mathrm{oc}}^{\mathrm{i}} \qquad [15]$$

in which f_{oc} is the organic carbon fraction of the solid phase (kilograms per kilogram) and $K_{\mathrm{oc}}^{\mathrm{i}}$ is the organic carbon partitioning coefficient (meters cubed per kilogram) of NAPL component i, which can be estimated from solubility data and molecular structure.

The partitioning of NAPL components between the gas phase and the NAPL, and between the gas phase and the aqueous phase, are also commonly modeled using instantaneous linear partitioning. The Raoult law is utilized to describe the partitioning of each NAPL component between the gas phase and the NAPL when the NAPL consists of multiple components. A different partitioning coefficient is needed for each NAPL component:

$$C_{\mathrm{g}}^{\mathrm{i}} = K_{\mathrm{go}}^{\mathrm{i}} C_{\mathrm{o}}^{\mathrm{i}} \qquad [16]$$

in which $C_{\mathrm{g}}^{\mathrm{i}}$ is the concentration of NAPL component i in the gas phase (kilograms per cubic meter), $K_{\mathrm{go}}^{\mathrm{i}}$ is the linear gas–NAPL partitioning coefficient (cubic meters per cubic meter) for NAPL component i, and $C_{\mathrm{o}}^{\mathrm{i}}$ is the concentration of NAPL component i in the NAPL (kilograms per cubic meter). Similarly, the Henry law is used to describe the partitioning of NAPL mass between the gas and aqueous phases:

$$C_{\mathrm{g}}^{\mathrm{i}} = K_{\mathrm{gaq}}^{\mathrm{i}} C_{\mathrm{aq}}^{\mathrm{i}} \qquad [17]$$

in which K^{i}_{gaq} is the linear gas–aqueous phase partitioning coefficient (cubic meter per cubic meter) for NAPL component *i*. The concentrations of the NAPL components in the gas phase can be converted to partial pressures, and vice versa, using either the ideal or some nonideal gas law. The Henry law and Raoult law appear to be the same. Both describe linear relationships for idealized gas–liquid partitioning; however, there is a difference. The proportionality constant in the Raoult law is the partial pressure of a solvent component in its pure form, while the proportionality constant in the Henry law depends on the standard state of a solute in a mixture, generally at infinite dilution. Consequently, ideal Henry law behavior is observed for low component concentrations, while ideal Raoult law applies to a component that is nearly pure (high concentration). The two are thermodynamically consistent; however, significant deviations are frequently observed at intermediate concentration (mole fraction) ranges.

Linear partitioning coefficients are also used to describe the partitioning of NAPL mass between the NAPL and the aqueous phase:

$$C^{i}_{aq} = K^{i}_{aqo} C^{i}_{o} \qquad [18]$$

in which K^{i}_{aqo} is the linear NAPL–aqueous phase partitioning coefficient (cubic meters per cubic meter) for NAPL component *i*. Because it is common to assume that C^{i}_{aq} will be at its equilibrium solubility concentration, eqn [18] can be rewritten as:

$$C^{i}_{aq} = K^{i}_{aqo} \chi^{i}_{o} \rho_{o} \qquad [19]$$

in which χ^{i}_{o} is the mole fraction of NAPL component *i* in the NAPL and ρ_o is the mass density of the NAPL (kilograms per cubic meter). As NAPL components partition to fluid and solid phases at different rates, ρ_o changes with time.

If mass transfer cannot be assumed to be instantaneous, rate-limited mass transfer needs to be used to describe changes in NAPL component partitioning. Rate-limited interphase mass transfer is often approximated by first-order kinetic relationships representing the combined effect of all physical nonequilibrium processes affecting the migration between phases. The driving force is the concentration gradient across the interfaces separating the phases.

In nonequilibrium first-order mass-transfer relationships across gas–water and water–solid interfaces, the interphase exchange term, I_{aqo} (kilograms per cubic meter per second) for first-order dissolution of NAPL into the aqueous phase is usually given by:

$$I^{i}_{aqo} = k^{i}_{aqo} a_{aqo} \left(\chi^{i}_{o} \rho_{o} - C^{i}_{aq} \right) = k^{i*}_{aqo} \left(\chi^{i}_{o} \rho_{o} - C^{i}_{aq} \right) \qquad [20]$$

where k_{aqo} (meters per second) and k^{*}_{aqo} (per second) are mass-transfer coefficients between the aqueous phase and NAPL, and a_{aqo} (per meter) is the specific interfacial area separating the phases. As shown in eqn [20], NAPL dissolution rates are dependent on interfacial contact area. Due to difficulties in obtaining interfacial areas experimentally, this parameter is usually not explicitly obtained, but implicitly contained in k^{*}_{aqo}. The relation between rate-limited NAPL dissolution and interfacial area has been the focus of several research projects.

Summary

The behavior of immiscible fluids (NAPLs) in soils is very complex, and there are some important physical and chemical processes that govern the fate of NAPLs and their components in soils. To predict the fate of NAPL components in soils, the convective, dispersive, and diffusive movement of the NAPL components, as well as their chemical and biological transformations and interphase mass transfer, must be described. The resulting mathematical equations are nonlinear partial-differential equations that cannot be solved analytically for general conditions. Therefore, numerical techniques are used to estimate the solutions. Besides difficulties in describing all of the relevant processes that affect the fate of NAPL in soils, there are difficulties in solving the complex system of mathematical equations for predicting NAPL fate in soils. Additional research, both experimental and theoretical, will be required to achieve a more complete understanding of these processes.

Acknowledgment

The above work has been authored, in part, by a contractor of the US government under DOE contract DE-AC07-99ID13727.

See also: **Capillarity**; **Hydrocarbons**; **Hydrodynamics in Soils**; **Hysteresis**

Further Reading

Adamson AW (1990) *Physical Chemistry of Surfaces*. New York: John Wiley.

Anderson WG (1987) Wettability literature survey. 4. Effects of wettability on capillary pressure. *Journal of Petroleum Technology* 39: 1283–1300.

Armstrong JE, Frind EO, and McClellan RD (1994) Nonequilibrium mass transfer between the vapor, aqueous, and solid phases in unsaturated soils during vapor extraction. *Water Resources Research* 30: 355–368.

Brooks RH and Corey AT (1966) Properties of porous media affecting fluid flow. *Journal of Irrigation and Drainage Division Proceedings of the ASCE* 92: 61–88.

Busby RD, Lenhard RJ, and Rolston DE (1995) An investigation of saturation-capillary pressure relations in two- and three-fluid systems for several NAPLs in different porous media. *Ground Water* 33: 570–578.

Corey AT (1994) *Mechanics of Immiscible Fluids in Porous Media*. Highlands Ranch, CO: Water Resources Publications.

Dane JH and Topp GC (2002) *Methods of Soil Analysis,* part 4, Physical Methods. Madison, WI: Soil Science Society of America.

Dane JH, Hofstee C, and Corey AT (1998) Simultaneous measurement of capillary pressure, saturation, and effective permeability of immiscible liquids in porous media. *Water Resources Research* 34: 3687–3692.

Dullien FAL (1992) *Porous Media: Fluid Transport and Pore Structure*. New York: Academic Press.

Fetter CW (1993) *Contaminant Hydrogeology.* New York: Macmillan.

Hillel D (1980) *Fundamentals of Soil Physics*. New York: Academic Press.

Hofstee C, Oostrom M, Dane JH, and Walker RW (1998) Infiltration and redistribution of perchloroethylene in partially saturated, stratified porous media. *Journal of Contaminant Hydrology* 34: 293–313.

Honarpour M, Koederitz L, and Harvey AH (1986) *Relative Permeability & Petroleum Reservoirs*. Boca Raton, FL: CRC Press.

Imhoff PT, Jaffe PR, and Pinder GF (1994) An experimental study of complete dissolution of a nonaqueous phase liquid in saturated porous media. *Water Resources Research* 30: 307–320.

Jury WA, Gardner WR, and Gardner WH (1991) *Soil Physics*. New York: John Wiley.

Lenhard RJ (1992) Measurement and modeling of three-phase saturation-pressure hysteresis. *Journal of Contaminant Hydrology* 9: 243–269.

Lenhard RJ (1994) Scaling fluid content–pressure relations of different fluid systems in porous media. In: Morel-Seytoux HJ (ed.) *Proceedings of 14th Annual Hydrology Days*, pp. 223–235. Atherton, CA: Hydrology Days Publications.

Lenhard RJ and Parker JC (1987) Measurement and prediction of saturation–pressure relationships in three-phase porous media systems. *Journal of Contaminant Hydrology* 1: 407–424.

Lenhard RJ and Parker JC (1987) A model for hysteretic constitutive relations governing multiphase flow. 2. Permeability–saturation relations. *Water Resources Research* 23: 2197–2206.

Lenhard RJ and Parker JC (1988) Experimental validation of the theory of extending two-phase saturation-pressure relations to three-fluid phase systems for monotonic drainage paths. *Water Resources Research* 24: 373–380.

Lenhard RJ, Parker JC, and Kaluarachchi JJ (1989) A model for hysteretic constitutive relations governing multiphase flow. 3. Refinements and numerical simulations. *Water Resources Research* 25: 1727–1736.

Mayer AS and Miller CT (1996) The influence of mass transfer characteristics and porous media heterogeneity on nonaqueous phase dissolution. *Water Resources Research* 32: 1551–1568.

Oostrom M and Lenhard RJ (1998) Comparison of relative permeability-saturation-pressure parametric models for infiltration and redistribution of a light nonaqueous-phase liquid in sandy porous media. *Advances in Water Resources* 21: 145–157.

Oostrom M, Hofstee C, Lenhard RJ, and Wietsma TW (2003) Flow behavior and residual saturation formation of liquid carbon tetrachloride in unsaturated heterogeneous porous media. *Journal of Contaminant Hydrology* 64: 93–112.

Parker JC and Lenhard RJ (1987) A model for hysteretic constitutive relations governing multiphase flow. 1. Saturation–pressure relations. *Water Resources Research* 23: 2187–2196.

Parker JC, Lenhard RJ, and Kuppusamy T (1987) A parametric model for constitutive properties governing multiphase flow in porous media. *Water Resources Research* 23: 618–624.

Powers SE, Abriola LM, and Weber WJ Jr (1994) An experimental investigation of nonaqueous phase liquid dissolution in saturated subsurface systems: transient mass transfer rates. *Water Resources Research* 30: 321–332.

Sposito G (1989) *The Chemistry of Soils*. New York: Oxford University Press.

Tiab D and Donaldson EC (1996) *Petrophysics*. Gulf Publishing.

van Genuchten MT (1980) A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Science Society of America Journal* 44: 892–898.

Willhite GP (1986) *Waterflooding*. Richardson, TX: Society of Petroleum Engineers.

INCEPTISOLS

A Palmer, Massey University, Palmerston North, New Zealand

Introduction

Inceptisols are found in a wide range of environments and include a variety of soils. They occur in cool to very warm and humid to subhumid climates, from the arctic to the tropics. In temperate areas with ustic or udic soil-moisture regimes, Inceptisols tend to be characterized by weak soil weathering, just sufficient for mild soil development. This is often expressed as an alteration of material from its original condition to the extent that silicate clays are formed, oxides are liberated, or soil structure has formed. Inceptisols often occur on more resistant quartz-rich deposits where the rate and degree of weathering are somewhat retarded. Elsewhere Inceptisols are merely soils that do not meet the criteria to be members of any of the other soil orders.

Inceptisols have also undergone major changes in classification since *Soil Taxonomy* was published. Soils formed in volcanic ash (formerly called Andepts) have now been grouped in the order Andisols, and less importance has been placed on soils with low organic matter content (formerly Ochrepts) compared with those with higher organic matter content in their topsoil (formerly Umbrepts). Now, more than ever, Inceptisols can be regarded as the 'sweepings' of *Soil Taxonomy.*

Occurrence

The main factors that determine the occurrence of Inceptisols are soil temperature, landscape position, and soil-moisture regime:

Soil Temperature Regime: Cryepts;
Landscape Position: Aquepts;
Soil-Moisture Regime: Xerepts (dry) – Ustepts – Udepts (moist).

Inceptisols are classified in the normal *Soil Taxonomy* manner of a cascade, where the aquic soil-moisture regime (poorly drained) takes precedence, followed by human disturbance, cold soil temperatures, and finally other soil-moisture regimes:

1. Permanent or intermittent high water table: Aquepts;
2. Profiles relating to human disturbance: Anthrepts;
3. Cold but not permanently frozen soil temperatures: Cryepts;
4. Dry season with soil-moisture deficit: Ustepts;
5. Pronounced soil-moisture deficit in summer: Xerepts;
6. Other Inceptisols that remain moist for most of the year: Udepts.

It is therefore convenient to discuss the Inceptisols under these headings.

Inceptisols with Permanent or Intermittent High Water Tables

Aquepts are the Inceptisols with a permanently or intermittently high water table. The water table may rise through the profile or be perched above an impermeable layer or pan when rainfall exceeds evapotranspiration. Inceptisols are common on alluvial plains, above normal flood level, where time has been adequate for formation of a cambic horizon. However, the Aquepts often occupy former back-swamp positions, or the toe positions of alluvial fans where water tables are high.

In temperate areas affected by land destabilization during cool climates of the late Pleistocene, loess is a common soil parent material on river terraces, fans, marine terraces, and gently sloping downlands. Particularly in areas with a ustic soil-moisture regime, where the loess is subject to annual wetting and drying cycles, fragipans form in the subsoil. Many of these loessial soils with fragipans develop perched water tables when rainfall is high and evapotranspiration rate is low (usually winter). The Inceptisols that develop are Fragiaquepts, but in these conditions soils also commonly exhibit clay translocation and formation of an argillic horizon to the extent that they fall into the Fragiaqualf category. On loess-covered marine terraces in the Manawatu region of New Zealand (Figure 1), the Tokomaru series may classify as either a Fragiaqualf or a Fragiaquept, because clay translocation in the Bgt horizon is borderline between the two orders.

However, these are not the only Aquepts in the Manawatu landscape. During the last glacial there was wholesale destabilization of the landscape. In particular the adjacent Mesozoic greywacke-argillite ranges (up to 1600 m) lost their forest cover; scree slopes fed huge quantities of detritus into river headwaters. The Tertiary and Pleistocene mudstone-sandstone hill country also became a grass-shrubland rather than forest, also resulting in wholesale erosion.

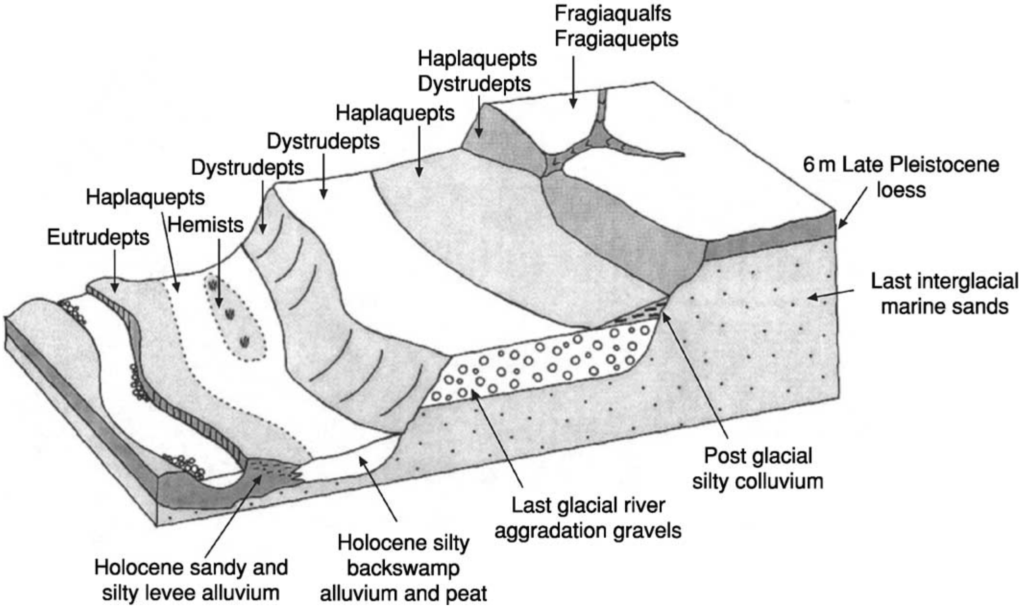


Figure 1 Soil-landscape diagram for part of the Manawatu region, southwest North Island, New Zealand. The rainfall is 1000 mm per annum with an ustic soil moisture regime. The upper terrace is comprised of Last Interglacial marine sands, subsequently covered with Last Glacial quartzo-feldspathic loess. In the prevailing climate Fragiaqualfs and Fragiaquepts form. The middle terrace is a river aggradation terrace from the last cooling of the Last Glacial. The inner side has been covered by postglacial colluvium in which Haplaquepts have formed, while Dystrudepts form on the stony gravel surface. On the lower (Holocene) alluvial terrace, Eutrudepts have formed on well-drained levees and Hemists and Haplaquepts on back-swamp positions.

As a result the local rivers became wide, braided beds and aggraded, forming large, low-angle fans that grade back high into the ranges. The main soils formed on the gravels and well-drained loamy alluvial cover are Dystrudepts.

However, as the climate warmed after 15 000 years BP forests recloaked the hill and mountain land, and the supply of gravels slowed. This led to the rivers downcutting, forming a series of degradation terraces. Preservation of the last glacial terrace was further enhanced by regional uplift. Fine-grained colluvium from adjacent terraces or hill country was washed on to part of the gravel terrace. Today, this colluvium is poorly drained, with a perched water table, and Haplaquepts have formed. On the Holocene degradation terraces, there is a repeated sequence of Eutrudepts on sandy textured levees, levee slopes, crevasse splays and point bars, and Haplaquepts and Hemists on back-swamp positions. On flood plains are Psamments, Orthents, and Aquepts.

All the Aquepts in this region pose problems for land use. They must be artificially drained but, even so, are difficult soils for horticulture and cropping. Loss of soil structure and formation of plough pans are the main problem. In pastoral farming, animal hooves and other traffic cause severe pugging (disturbance of the topsoil layer) in winter. Paradoxically, the finer-textured Aquepts are also drought prone in summer, because these soils have a pore-size distribution unsuited to storing adequate plant-available water.

Aquepts occupy similar back-swamp positions on many of the worlds major rivers, particularly in lower reaches; for example, Aquepts of the Indo-Ganges Plain and Aquepts of the lower Mississippi Valley.

Anthrepts

Anthrepts are the Inceptisols that are found in areas of prolonged cropping, manuring, and effluent disposal associated with human activities. They have either anthropic epipedons (Haplanthrepts) or plaggen epipedons (Plagganthrepts).

Anthropic epipedons result from prolonged irrigation and fertilizer application to produce crops, and from disposal of bones, shells, and wastes as a result of human occupation. The anthropic epipedon has the same organic matter content, thickness, structure and colour as mollic epipedons, but has higher phosphate content. Anthropic epipedons on irrigated land usually occur in arid areas, where application of water and fertilizer has produced an artificial epipedon that would not be there were it not for the water. Examples of these soils can be found in dry, irrigated areas of the USA (e.g., Idaho, California, Texas), Mexico, the Middle East, and Asia. Anthropic epipedons resulting from disposal of bones, shells, and wastes (middens) are found in many parts of the world but are very localized, e.g., North America from former Indian occupation, the Pacific Islands and New Zealand from Polynesian or Melanesian occupation and Europe.

Plaggen epipedons are produced when the manure and bedding materials (sawdust, straw, pine needles) from housed or penned animals is applied to local soils, resulting in overthickened topsoils on otherwise less-fertile soils.

Cryepts

Cryepts occur on the margins of tundra at high latitudes, and at high altitude on major mountain ranges and plateaux. Soils that have permafrost or gelic materials within the top 1 m are now called Gelisols. Therefore Cryepts are common in coastal areas near the Alaska Peninsula, in northern Saskatchewan, and Western Siberia, bordering Gelisols of the Arctic regions. Cryepts are also common on high mountain ranges such as the Rocky Mountains in west-central USA, the Urals, Himalayas, and other ranges and plateaux of Central Asia and the Andes.

Cryepts have formed mostly under coniferous or mixed coniferous hardwood forest. Parent materials are till, solifluction deposits, loess, colluvium, or bedrock regolith of late Pleistocene or Holocene age. It is most common for ochric epipedons to overlie cambic horizons.

Ustepts

Ustepts are the Inceptisols that receive a seasonal rainfall, are not poorly drained, and most commonly have ochric epipedons and cambic horizons. The discriminating criteria at great group level are the presence of duripans or petrocalcic horizons, and whether the parent material is leached or contains free calcium carbonate:

With a duripan: Durustepts;
With a calcic or petrocalcic horizon: Calciustepts;
Leached or noncalcareous parent materials: Dystrustepts;
Other Ustepts: Haplustepts.

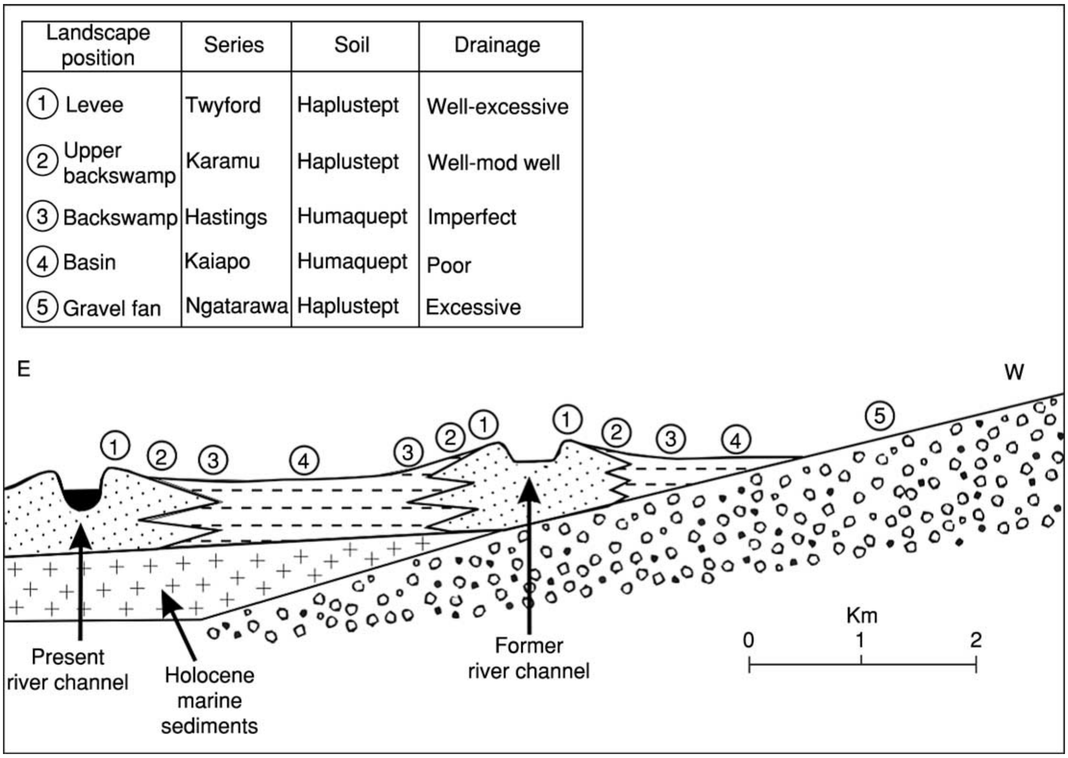
Soils with ustic soil-moisture regimes have commonly formed on early Holocene or Pleistocene parent materials, under native grasslands, or open forests near the grassland–forest transition. These areas can have had a complex history of forestation and deforestation during climate cycles of the Pleistocene and early Holocene, often assisted by fire during droughty periods. In the western USA particularly on the Great Plains, Ustepts commonly grade laterally into vast areas of Mollisols, where the history of grassland vegetation has been more uniform, and where stable forms of organic matter have been able to accumulate to greater depth. These Ustepts usually have umbric epipedons and might have truncated topsoils, not meeting the depth criteria for a mollic epipedon. In western Texas and New Mexico, there is a transition from Argids and Calcids in arid areas to Calciustepts and Ustolls as rainfall increases on the grassland Praries toward the east. On river terraces dating back to the early Pleistocene and Tertiary on the Great Plains in northern Texas and the Oklahoma panhandle, soils are formed in fine eolian and fluvial sediments. The dominant soils are Paleustolls and Argiustolls, with Psamments on the river flood plains. However, some younger terraces, and eroded soils or soils developed in reworked materials on the higher terraces are often Calciustepts.

In Hawkes Bay, on the east coast of the North Island of New Zealand, Inceptisols form on the low river terraces (Figure 2). The Twyford series (Haplustept) forms on levees and crevasse splays of current and abandoned river channels. The Karamu (Haplustepts) and Hastings series (Humaquepts) occupy backswamp positions, and Kaiapo series (Humaquepts) are found in the low-lying basins between levee systems. The Hastings Series is most extensive and has been deeply drained so that it is a most productive soil for pip- and stonefruit, vegetables, and other field crops. The Twyford and Karamu soils can be equally productive but require irrigation to ensure plant survival. Haplustepts of the Ngatarawa series are also found on older gravelly alluvium that dates back to the early Holocene and late Pleistocene. The most extensive land use on these soils is for vineyards.

Xerepts

These soils occur in areas with a ‘Mediterranean climate,’ typified by moist and cool winters, and warm dry summers. The winter precipitation is effective, because it coincides with minimum evapotranspiration rates. They are usually well drained and occur in frigid to thermic soil temperature regimes. Most have ochric or umbric epipedons and cambic horizons, but duripans, fragipans, and calcic horizons occur in some great groups:

Xerepts with duripans: Durixerepts;
Other Xerepts with calcic or petrocalcic horizons: Calcixerepts;
Xerepts with fragipans: Fragixerepts;
Leached Xerepts with moderate or low base saturation: Dystroxerepts;
Other Xerepts: Haploxerepts.



Landscape position	Series	Soil	Drainage
① Levee	Twyford	Haplustept	Well-excessive
② Upper backswamp	Karamu	Haplustept	Well-mod well
③ Backswamp	Hastings	Humaquept	Imperfect
④ Basin	Kaiapo	Humaquept	Poor
⑤ Gravel fan	Ngatarawa	Haplustept	Excessive

Figure 2 A cross-section of part of the Heretaunga plains on the east coast of the North Island of New Zealand. The rainfall is 780 mm with an ustic soil moisture regime. The plains were an embayment of the sea until being infilled by alluvium in the last 2000 years. Late Last Glacial to early Holocene gravels are overlain by transgressive marine sediments and then by the river alluvium. A complicated pattern of levees, backswamps, and interchannel basins has formed. The soils formed relate closely to geomorphology.

Xerepts are common, as the climate connotation suggests, in Mediterranean countries, particularly Spain, Italy, Turkey, and northern Tunisia, Algeria, and Morocco. In the USA Xerepts are most common in the western states of California, Oregon, Washington, Idaho, and Utah. Open mixed coniferous forest and grassland dominates in cooler climate, whereas mixed oak, shrub, and grassland is more common in thermic temperature regimes.

Udepts

Udepts are the Inceptisols that remain moist for most of the year. Dry periods are not long enough in most years to have soil-moisture deficits for as long as 3 months, i.e., they occur in a udic soil-moisture regime. Most have developed under a forest or forest–shrubland grassland vegetation. They are usually well or moderately well drained, with ochric or umbric epipedons over cambic horizons. The various great groups are named for the presence of other horizons:

Presence of a near-surface sulfuric horizon: Sulfudepts;
Presence of a duripan: Durudepts;
Presence of a fragipan: Fragiudepts;
Free carbonates or high base status: Eutrudepts;
Remaining Udepts: Dystrudepts.

Dystrudepts and Fragiudepts are common in the northeastern USA, on non-calcareous glacial till in Pensylvania and New York, and on acid shales and sandstones in the Appalachian Mountains and Allegheney Plateau. In forested areas, on acid or coarse-textured parent materials, Dystrudepts commonly form a mosaic with Spodosols. On the west coast of the South Island of New Zealand, in the Smoky Mountains, at the southern end of the Appalachian mountains, there is a similar mosaic of Spodosols and Drystrudepts. Both soils are deeply weathered, as indicated by destruction of feldspars and formation of gibbsite, kandite, and vermiculite-chlorite in the soil B horizon. The only observable difference is that Spodosols are more frequent on coarse-textured but less well drained sites, with slightly different flora.

The soils that form often relate to variations in parent material. In the soils formed on Wisconsin till in New Jersey, Dystrudepts have been developed on till derived from gneissic rocks, whereas adjacent soils formed in quartzose till and calcareous till are Spodosols and Alfisols, respectively. There has been considerable translocation of materials in the Spodosols and Alfisols; whereas, in the iron-rich gneissic material, iron release has been gradual and the residual iron has coated local resistant particles.

Inceptisols may also occur in landscapes dominated by other orders representing strongly weathered soils, by occurring on less-stable positions in the landscape. In the Carolinas, Dystrudepts occur on steep slopes in the granite-gneiss terrain above the fall line. On more stable parts of the granite-gneiss terrain are Udalfs and Udults, while on coastal plain sediments below the fall line are Psamments, Spodosols, Udults, and Aquults. At the southern end of the Appalachians, in Tennessee, Dystrudepts occur on upland positions on shale, where mass wasting, overland flow, and soil creep inhibit soil formation. On adjacent, stable shoulder slopes, Hapludults occur, reflecting time enough to form a strongly leached Bt rather than a Bw horizon. On toeslope and floodplain bottom lands are Endoaquepts, while Hapludalfs occur on alluvial terraces.

In southeastern Brazil on steep, pre-Paleozoic gneiss and granite, deep Hapludox is found on stable summits, Dystrudepts on shoulder slopes, shallow Hapludox on backslopes, Kandiudults on footslopes, and Eutrudepts on toeslopes. The shoulder slope and backslope positions in this landscape are considered to be unstable retreating slopes, enabling soil rejuvenation. Similarly, a little farther south, Inceptisols are found on a retreating escarpment of Cretaceous subarkosic sandstone, whereas Oxisols occur on old fluvial terraces and pediments below the escarpment, and on the stable summit above the escarpment. Shoulder slopes and footslopes are occupied by Alfisols and Ultisols.

In the southern South Island of New Zealand, a climosequence of Inceptisols occurs in late Pleistocene loess deposits (Figure 3). As rainfall increases from 600 to 1200 mm, the soils change from Lamellic Haplustepts (600 mm) to Aeric Fragiaquepts and Aquic Haplustepts (700 mm); Aeric Fragiaquepts and Aquic Fragiudepts (850 mm); Aeric Fragiaqupts (950 mm); Aquic Dystrudepts (1100 mm); and Typic Dystrudepts (1200 mm). A feature of this sequence is first the deterioration then improvement of drainage as rainfall increases. The reason lies in the development of a gammate fragipan, which causes winter moisture to perch in the subsoil above, and then its disappearance at higher rainfall. Profiles above 1000 mm do not dry out substantially in most summers, whereas profiles in the range 700–1000 mm experience strong fluctuations of soil moisture, with winter surpluses and summer deficits.

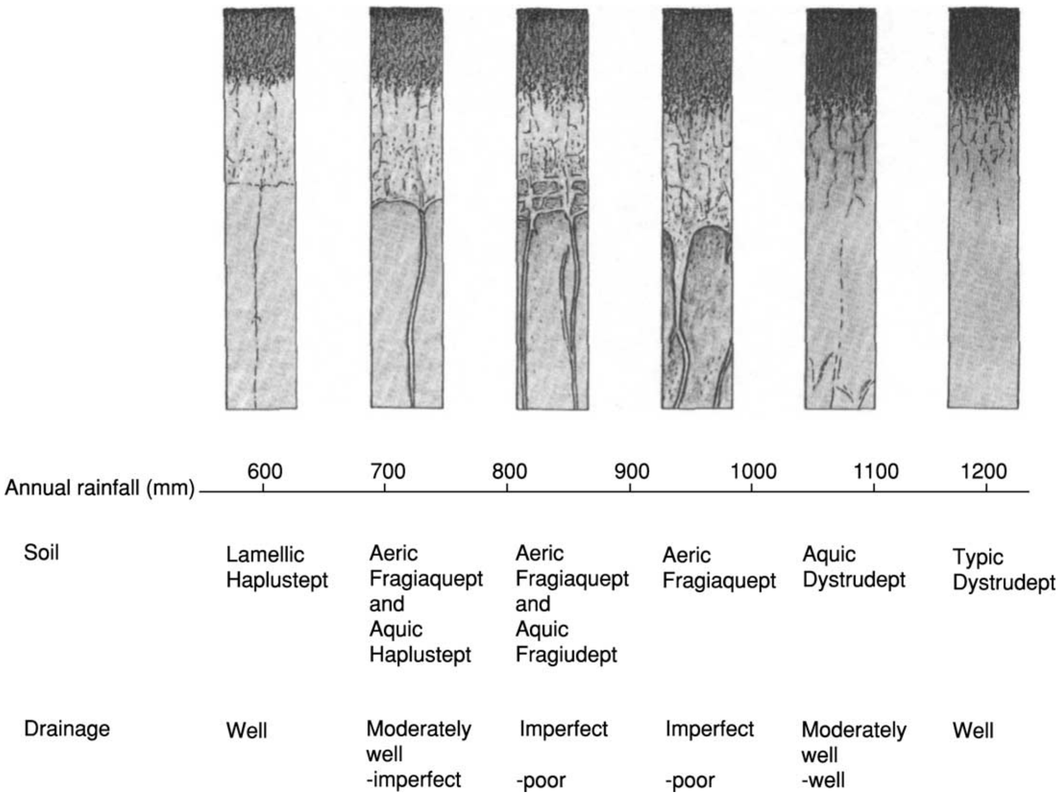


Figure 3 A climosequence of soils developed in Last Glacial loess in Southland, New Zealand. As rainfall increases, the drainage first deteriorates and then improves, influenced by the formation of an impermeable gammate fragipan at intermediate rainfalls (adapted with permission of the NZSSS).

At Otaki in the southwestern North Island of New Zealand, at a rainfall of approximately 1050 mm, a complex pattern of poorly to well-drained soils occurs on late last glacial and early Holocene loamy alluvium. The soils occur as a complex on gently undulating, last glacial river aggradation terraces. Soil series are defined according to their drainage status. The soils are Endoaquepts (poorly drained); Eutrudepts (imperfectly drained); and moderately well and well-drained Dystrudepts. Various reasons can be proposed for this soil pattern, but none have been proven to date. The soil pattern has a grain or orientation that suggests an influence of fluvial deposition, but no relationship has been established with either texture, depth to underlying gravels, or landscape position.

The Dystrudepts often have a trace of allophane, and the suggestion has been made that these soils contain more volcanic ash than adjacent soils, because allophane is a recognized weathering product of volcanic glass. Soils rich in allophane have good structure and are generally well drained. The drainage status is reflected in the chemical and physical properties of the soils, with well-drained soils having lower bulk density, higher macroporosity, and lower pH and base saturation than poorly drained soils. The well-drained soil also has less mica clay, more chlorite, hydroxy-interlayered vermiculite instead of vermiculite, and less kandite than the poorly drained soil.

Another similar soil complex of poorly drained to well-drained soils occurs in the same district on flat to gently undulating loess-covered terraces. In this complex, the soils are: poorly drained Typic Haplaquepts; imperfectly drained Typic Haplaquepts; moderately well drained Typic Dystrudepts, and well drained Umbric Dystrudepts. The latter soil contains appreciable (up to 10%) allophane in its upper horizons, but not enough to be an Andisol. Available mineralogy suggests that the loess was initially uniform, dominated by quartzo-feldspathic components, but with a small tephric component. Early in the weathering process, a critical threshold was reached, where in some locations, the volcanic glass formed allophane, while in other places formed halloysite. Where allophane formed, leaching of silica continued, more allophane formed, the allophane promoted stable, subangular blocky, granule and crumb structure, moderate macroporosity, and well-drained profiles. Where halloysite formed, leaching of silica was inhibited, blocky structure formed in a less-porous, denser soil. The soil distribution and soil properties suggest that with time conditions have more and more favored weathering to halloysite clay, and formation of Haplaquepts rather than Dystrudepts.

The described soil complexes occur where the rainfall is between 1050 and 1150 mm. At lower rainfall, the same alluvium and loess weathers to Fragiaqualfs, Fragiudepts, and Haplaquepts, while, at higher rainfall, Dystrudepts and Hapludands dominate, i.e., soils receiving higher rainfall have better drainage than those with lower rainfall.

See also: **Classification Systems:** USA; **Pedology:** Basic Principles

Further Reading

Almond PC and Tonkin PJ (1999) Pedogenesis by upbuilding in an extreme leaching and weathering environment, and slow loess accretion, south Westland, New Zealand. *Geoderma* 92: 1–36.

Anjos LH, Fernandes MR, Peteira MG, and Franzmeier DP (1998) Landscape and pedogenesis of an oxisol-inceptisol-ultisol sequence in southeastern Brazil. *Soil Science Society of America Journal* 62: 1651–1658.

Aslan A and Autin WJ (1998) Holocene flood-plain soil formation in the southern lower Mississippi Valley, implications for interpreting alluvial paleosols. *Geological Society of America Bulletin* 110: 433–449.

Blume HP and Kalk E (1986) Bronzezeitlicher Auftragsboden bei Rantum auf Sylt. *Zeitschrift fur Pflanzenernahrung und Bodenkunde* 149: 608–613.

Bruce JG (1983) Effect of climate on morphological features of soils from loess in the southern part of South Island, New Zealand. *Australian Journal of Soil Research* 21: 359–371.

Buol SW, Hole FD, and McCracken RJ (1980) *Soil Genesis and Classification*. Ames: Iowa State University Press.

Carter BJ, Ward PA, and Shannon JT (1990) Soil and geomorphic evolution within the Rolling Red Plains using Pleistocene volcanic ash deposits. *Geomorphology* 3: 471–488.

Dipak Sarkar Sahoo AK (2000) Aquepts of the Indo-Gangetic Plain of Bihar and their suitability for some major crops. *Journal of the Indian Society of Soil Science* 48: 561–566.

Duncan MM and Franzmeier DP (1999) Role of free silicon, aluminum, and iron in fragipan formation. *Soil Science Society of America Journal* 63: 923–929.

Engel RJ, Witty JE, and Eswaran H (1997) The classification, distribution, and extent of soils with a xeric moisture regime in the United States. *Catena* 28: 203–209.

Foss JE, Moormann FR, and Rieger S (1983) Inceptisols. In: Wilding LP, Smeck NE, and Hall GF (eds) *Pedogenesis and Soil Taxonomy. II. The Soil Orders*. Amsterdam: Elsevier.

Holliday VT (1990) Soils and landscape evolution of eolian plains: the southern High Plains of Texas and New Mexico. *Geomorphology* 3: 489–515.

Karlstrom ET (2000) Use of soils to identify glacial deposits of various ages east of Glacier National Park, Montana, USA. *Arctic Antarctic and Alpine Research* 32: 179–188.

Krebs RD and Tedrow JFC (1957) Genesis of three soils derived from Wisconsin till in New Jersey. *Soil Science* 83: 207–218.

Lepsch IF, Buol SW, and Daniels RB (1977) Soil-landscape relationships in Occidental Plateau of Sao Paulo State, Brazil: I. Geomorphic surfaces and soil mapping units. *Soil Science Society of America Journal* 41: 104–109.

Lepsch IF, Buol SW, and Daniels RB (1977) Soil-landscape relationships in Occidental Plateau of Sao Paulo State, Brazil: II. Soil morphology, genesis, and classification. *Soil Science Society of America Journal* 41: 109–115.

McCracken RJ, Shanks RE, and Clebsch EEB (1962) Soil morphology and genesis at higher elevations of the Great Smokey Mountains. *Soil Society of America Proceedings* 26: 384–388.

McFadgen BG (1980) Maori plaggen soils of New Zealand, their origin and properties. *Journal of the Royal Society of New Zealand* 10: 3–19.

Molloy L (1998) *Soils in the New Zealand Landscape: The Living Mantle*. Lincoln: New Zealand Society of Soil Science.

Pape JC (1970) Plaggen soils in the Netherlands. *Geoderma* 4: 229–255.

Phillips DH, Foss JE, Stiles CA, Trettin CC, and Luxmoore RJ (2001) Soil–landscape relationships at the lower reaches of a watershed at Bear Creek: near Oak Ridge Tennessee. *Catena* 44: 205–222.

Pillans B, McGlone M, Palmer AS, Mildenhall D, Alloway BV, and Berger G (1993) The last Glacial maximum in central and southern North Island, New Zealand: a paleoenvironmental reconstruction using the Kawakawa Tephra Formation as a chronostratigraphic marker. *Palaeogeography, Palaeoclimatology, Palaeoecology* 101: 283–304.

Ruhe RV and Olson CG (1980) Clay-mineralogy indicators of glacial and non-glacial sources of Wisconsinan loesses in southern Indiana. *Geoderma* 24: 283–297.

Soil Survey Staff (1975) *Soil Taxonomy: A Basic System of Soil Classification for Making and Interpreting Soil Surveys*. Agriculture Handbook. United States Department of Agriculture. Washington: US Government Printing Office.

Soil Survey Staff (1999) *Soil Taxonomy: A Basic System of Soil Classification for Making and Interpreting Soil Surveys*, 2nd ed. Agriculture Handbook 436. United States Department of Agriculture. Washington: US Government Printing Office.

Ugolini FC and Schlichte AK (1973) The effect of Holocene environmental changes on selected western Washington soils. *Soil Science* 116: 218–227.

Viereck LA, Dyrness CT, Cleve K, and van Foote MJ (1983) Vegetation, soils, and forest productivity in selected forest types in interior Alaska. *Canadian Journal of Forest Research* 13: 703–720.

Yaalon DH (1997) Soils in the Mediterranean region: what makes them different? *Catena* 28: 157–169.

Industrial Pollution *See* **Pollution:** Industrial

INFILTRATION

T P A Ferré and A W Warrick, University of Arizona, Tucson, AZ, USA

Introduction

Infiltration is defined as the flow of water from aboveground into the subsurface. The topic of infiltration has received a great deal of attention because of its importance to topics as widely ranging as irrigation, contaminant transport, groundwater recharge, and ecosystem viability. More generally, a quantitative understanding of this process is vital to our ability to relate surface and subsurface processes in describing the hydrologic cycle.

While the definition of infiltration is simple, it can involve all aspects of flow through a variably saturated porous medium, ranging in complexity from steady-state, saturated flow in a homogeneous, isotropic medium to transient, unstable, unsaturated flow through an anisotropic, heterogeneous medium. The rate and pattern of infiltration vary with the distribution and rate with which water is supplied at the ground surface, the depth of the water table, the hydraulic properties of the subsurface materials, and the antecedent moisture content distribution. The following discussion focuses on infiltration into a homogeneous, isotropic medium as a basis for understanding the infiltration process.

Plug Flow

Constant Infiltration at the Ground Surface

In its simplest form, infiltration can be visualized as a one-dimensional, downward displacement process. Water application at the ground surface can be considered in two general ways. The simplest ground-surface boundary condition to envision, physically, is the application of water at a constant rate, such as infiltration during a constant-intensity rainfall. This is known as a type II, Neumann boundary condition. A type I, Dirichlet boundary condition, describes the application of water at a constant pressure head (or water content).

For simplicity, we begin by considering water supplied at a constant rate at the ground surface (type II ground-surface boundary condition). This rate can be defined as a flux density, with units of length per time, and is equivalent to the flow of water (volume per time) over an area. If the rate of supply is greater than the saturated hydraulic conductivity of the medium, water will eventually pond at the ground surface and flow overland. If the rate of water delivery is lower than the saturated hydraulic conductivity, all of the water will infiltrate. If the spatial area over which water is supplied is relatively large, infiltration can be approximated as one-dimensional downward. The water content of the medium will increase until the hydraulic conductivity of the medium at that water content is equal to the rate of water supply and infiltration will proceed under a unit gradient condition (i.e., the magnitude of the hydraulic head gradient is 1). As water is continually supplied, this wetted region will advance downward. Assuming that the water velocity in front of the advancing water is nearly zero, which is reasonable unless the antecedent water content is quite high, the added water will advance as a sharp wetting front. The depth to the wetting front, z_f, after a time t, can be determined from the antecedent volumetric water content, θ_i, the water content behind the wetting front, θ_wet, and the flux density, q, by a simple mass balance calculation:

$$z_\mathrm{f} = \frac{qt}{\theta_\mathrm{wet} - \theta_\mathrm{i}} \qquad [1]$$

The rate of advance of the wetting front, v_f, can be determined as the derivative of z_f with time:

$$v_\mathrm{f} = \frac{q}{\theta_\mathrm{wet} - \theta_\mathrm{i}} \qquad [2]$$

Notice that if the initial water content and the water content behind the wetting front are constant with depth and, given that the applied flux is constant in time, the velocity of the wetting front is constant with time. The rate of advance of the wetting front increases with an increase in the flux or the antecedent water content and with a decrease in the water content behind the wetting front.

As a first approximation, infiltrating water is generally assumed to displace the antecedent resident pore water completely. For example, consider that the water that is present in the medium before infiltration begins is dyed blue and that the infiltrating water is dyed red. There is a transition from a higher water content to a lower water content at a depth of z_f; but the water on either side of the wetting front is blue. At the same time, t, there is a transition from red water to blue water at a depth, z'_f, that is equal to:

$$z'_\mathrm{f} = \frac{qt}{\theta_\mathrm{wet}} \qquad [3]$$

The rate of advance of the applied water front, v'_f, is:

$$v'_\mathrm{f} = \frac{q}{\theta_\mathrm{wet}} \qquad [4]$$

Notice that if the water content behind the wetting front is constant with depth and the applied flux is constant in time, the velocity of the wetting front is also constant with time. The rate of advance of the applied water is independent of the antecedent water content. This simple representation of infiltration can be extended to consider flow through a layered subsurface in which each layer can have a different value for θ_i and θ_wet.

Constant Pressure Head at the Ground Surface

Now we consider the advance of a wetting front in response to a type I ground-surface boundary condition. As shown above, the velocity of the wetting front can be expressed as:

$$\frac{\mathrm{d}z_\mathrm{f}}{\mathrm{d}t} = v_\mathrm{f} = \frac{q}{\theta_\mathrm{wet} - \theta_\mathrm{i}} \qquad [5]$$

To consider the case of the advance of the wetting front in response to a constant pressure head applied at the ground surface, ψ_0, the pressure head at the wetting front is defined as ψ_f. Then the flux across the ground surface is:

$$q = -K_\mathrm{wet}\frac{\psi_\mathrm{f} - \psi_0 - z_\mathrm{f}}{z_\mathrm{f}} \qquad [6]$$

where K_wet is the hydraulic conductivity of the medium behind the wetting front. Substituting for the flux gives:

$$v_\mathrm{f} = \frac{\mathrm{d}z_\mathrm{f}}{\mathrm{d}t} = -K_\mathrm{wet}\frac{\psi_\mathrm{f} - \psi_0 - z_\mathrm{f}}{z_\mathrm{f}(\theta_\mathrm{wet} - \theta_\mathrm{i})} \qquad [7]$$

Integrating this expression gives an expression for the time, t, for the wetting front to advance to a depth z_f as:

$$t = \frac{(\theta_{\mathrm{wet}} - \theta_{\mathrm{i}})}{K_{\mathrm{wet}}}\left(z_{\mathrm{f}} + (\psi_{\mathrm{f}} - \psi_0)\ln\left(1 + \frac{z_{\mathrm{f}}}{\psi_0 - \psi_{\mathrm{f}}}\right)\right) \quad [8]$$

Green and Ampt derived a similar expression. For short times, gravity can be ignored and the results simplify to:

$$z_{\mathrm{f}} = \left[\frac{2(\psi_0 - \psi_{\mathrm{f}})K_{\mathrm{wet}}}{\theta_{\mathrm{wet}} - \theta_{\mathrm{i}}}\right]^{0.5} t^{0.5} \quad [9]$$

The rate of advance of the wetting front at time t is described as the derivative of z_f with t:

$$v_{\mathrm{f}} = 0.5\left[\frac{2(\psi_0 - \psi_{\mathrm{f}})K_{\mathrm{wet}}}{\theta_{\mathrm{wet}} - \theta_{\mathrm{i}}}\right]^{0.5} t^{-0.5} \quad [10]$$

Notice that, unlike the advance of a wetting front in response to a constant applied flux, the rate of advance of the wetting front in response to a type I ground-surface boundary condition changes with time. The rate of advance of the wetting front is highest initially when the distance from the surface to the wetting front is smallest, leading to the highest pressure-head gradient. This indicates that the rate of infiltration at the ground surface is highest at early times. At long times, gravity dominates infiltration and the infiltration rate at the ground surface decreases to a constant rate, q. As a result, at later times, the rate of advance of the wetting front decreases to $q/(\theta_{\mathrm{wet}} - \theta_{\mathrm{i}})$. That is, for larger times the wetting front in response to a type I ground-surface boundary condition advances at the same rate as a wetting front in response to a type II boundary condition with an equivalent applied flux.

Richards Equation – The Shape of the Wetting Front

The preceding treatments are sufficient to provide simplified descriptions of the one-dimensional advance of a wetting front into a homogeneous, drained soil under time-invariant boundary conditions. However, both of these analyses are based on the assumption of a step change in water content at the wetting front. A more complete, detailed description of the infiltration process requires solution of the flow equation, including constitutive relationships describing the water content–pressure head relationship for the soils into which water is infiltrating.

In general, the Richards equation is a nonlinear, parabolic partial differential equation that describes water movement through a variably saturated medium under the assumption of zero air resistance. The equation and the accompanying constitutive relationships can be written in many forms. For example, the mixed form describing one-dimensional downward water flow based on the water content, θ, and pressure head, ψ, is:

$$\frac{\partial\theta}{\partial t} = \frac{\partial}{\partial z}\left[K(\theta)\left[\frac{\partial\psi}{\partial z} + 1\right]\right] \quad [11]$$

In the preceding examples, the change in the volumetric water content was abrupt, taking one of two values, θ_{wet} or θ_{i}. In reality, the water content can assume any value between the residual water content and the porosity; the water content is a function of the capillary pressure and the antecedent conditions (hysteresis). If the water content varies smoothly across the wetting front, then the hydraulic conductivity will also take a continuous range of values. The nature of the dependence of the hydraulic conductivity on the water content controls the deviation of the shape of the wetting front from the idealized sharp boundary assumed above.

The following sections show examples of the infiltration process calculated using HYDRUS-2D, one of several available numerical models that solve the Richards equation for a range of boundary conditions and soil properties. The hydraulic properties for silt and sand were taken from the soil catalogue available through ROSETTA, which is included in HYDRUS-2D. These cases have been chosen to illustrate the influences of soil properties and water application rates on infiltration.

The Advance of a Wetting Front

For one-dimensional infiltration into a soil profile overlying a deep water table in response to a constant applied flux at the ground surface, the water content and the water pressure at the ground surface increase from their initial values to constant, higher values. These values are equal to the water content and pressure head at which the hydraulic conductivity of the soil equals the infiltration rate. The wetting front remains sharp because the hydraulic conductivity of the relatively dry medium below the wetting front is much lower than the hydraulic conductivity above the wetting front (solid lines in Figure 1a). As a result, a large gradient is needed across the wetting front to drive water into the soil. The pressure heads correspond, having the general shape of the solid lines in Figure 1b. The exact corresponding values of pressure head and water contents will depend on the soil-water characteristic curve, i.e., the relationship of θ to ψ.

For one-dimensional infiltration into a soil profile overlying a deep water table in response to a constant, increased pressure head at the ground surface, the water content and the water pressure at the ground-surface increase from their initial values to constant, higher values. For comparison, consider a ground surface boundary condition that corresponds with the constant infiltration rate applied in creating Figure 1. As for a type II ground-surface boundary condition, the wetting front remains sharp because the hydraulic conductivity of the relatively dry medium below the wetting front is much lower than the hydraulic conductivity above the wetting front (dashed lines in Figure 1a). As a result, a large gradient is needed across the wetting front to drive water into the soil. The pressure heads correspond, having the general shape of the dashed lines in Figure 1b. The exact corresponding values of pressure head and water contents depend on the soil-water characteristic curve, i.e., the relationship of θ to ψ. Notice that the wetting front has advanced to a greater depth at any given time for the constant pressure head ground-surface boundary condition than for the corresponding constant flux boundary condition. This is due to an early high pressure head gradient, leading to high initial fluxes under the constant-head boundary condition. At late times, the wetting front advances at the same rate for both boundary conditions, leading to a constant distance between the solid and dashed lines on Figure 1.

The Effects of Infiltration Rate

For simplicity, the remaining examples will only consider infiltration in response to a constant applied flux. For these conditions, for a given soil, the water content behind the wetting front increases with increases in the rate of infiltration until full saturation is reached. The wetting front velocity has been shown (above) to advance at a rate that is proportional to the applied flux and inversely proportional to the water content behind the wetting front. The rate of advance of the wetting front increases with higher rates of water application. However, note that the difference in water pressure at the surface (Figure 2b) is small compared with the difference in water contents (Figure 2a) for the range of water application rates chosen.

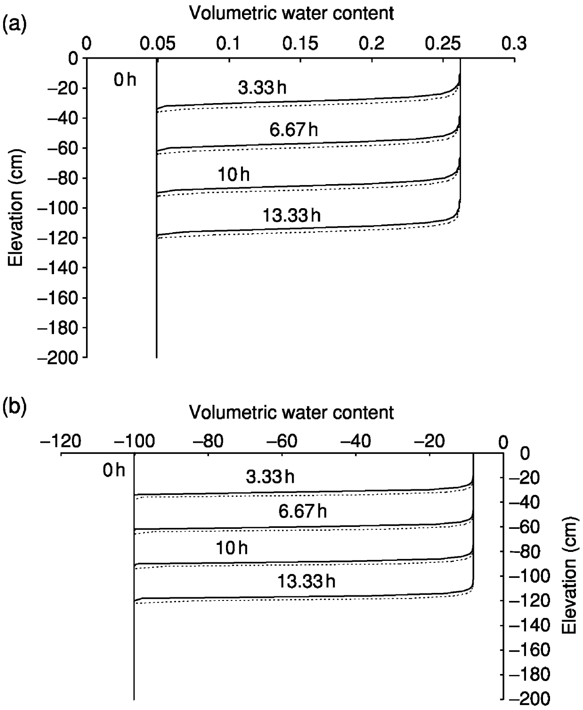


Figure 1 (a) Water-content and (b) pressure-head profiles after 3.3, 6.7, 10, 13.3, and 16.7 h of infiltration into a drained sand (−100-cm pressure head throughout). Solid lines show the response to a type II ground-surface boundary condition with a constant applied flux of 0.0005 cm s^{-1}. Dashed lines show the response to a type I ground-surface boundary condition with a constant applied pressure head of −8.08 cm.

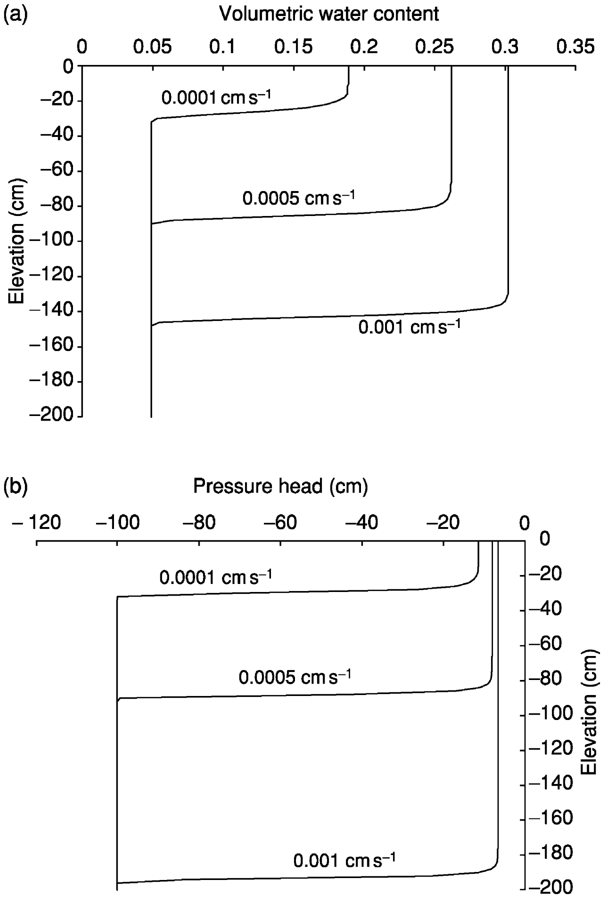


Figure 2 (a) Water-content and (b) pressure-head profiles after 10 h of infiltration into a drained sand at rates of 0.001, 0.0005, or 0.0001 cm s^{-1}.

The Effects of the Antecedent Water Content

As infiltration progresses, the downward-moving water continually displaces preexisting pore water and fills drained pores. For a sufficiently long profile that has no flow across the bottom boundary, the total volume of water added is equal to the total change in water stored in the profile. If the antecedent water content is high, the change in storage per unit length is small (Figure 3). As a result, the wetting front will progress deeper for a given total volume of added water.

The Effects of the Saturated Hydraulic Conductivity

If one medium has a higher saturated hydraulic conductivity than another and all other hydraulic properties are the same, then the first soil is able to transmit a given flux at a lower water content than the second soil (Figure 4). As a result, for a given rate of water application, the water content behind the wetting front is lower for the first, more conductive soil. To accommodate the same total change in water storage throughout the profile in response to a given volume of applied water, the wetting front moves deeper in the more conductive soil.

The Effects of Soil Texture

The texture of a soil strongly influences its hydraulic properties. For example, at low pressure heads, a fine-textured material tends to have a higher water content than a more coarse material. As a result, the antecedent water content may be higher in a silt or clay than in a sand. This can lead to a fine-textured soil having a higher hydraulic conductivity and a lower storage capacity than a coarse-textured soil for the same initial pressure head (Figure 5). At higher water contents, a coarse soil has a higher hydraulic conductivity than a fine soil. Therefore, at a given rate of water application, the water content behind the wetting front is lower for a sand than for a silt or clay. The balance of these effects for a given rate of water application and for given initial conditions determines the relative rate of advance of a wetting front through any two soils.

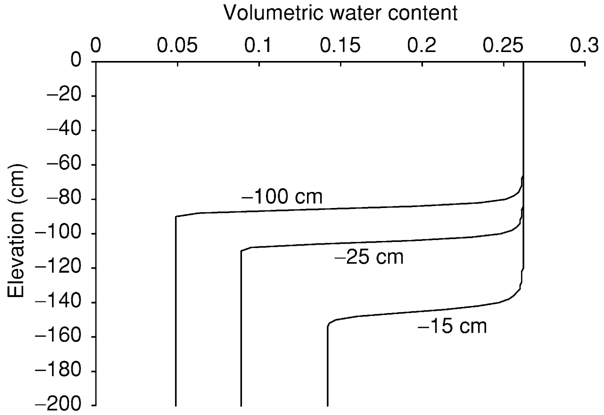


Figure 3 Water-content profiles after 10 h of infiltration at a rate of 0.0005 cm s^{-1} into a sand with an initial pressure head of −15, −25, or −100 cm.

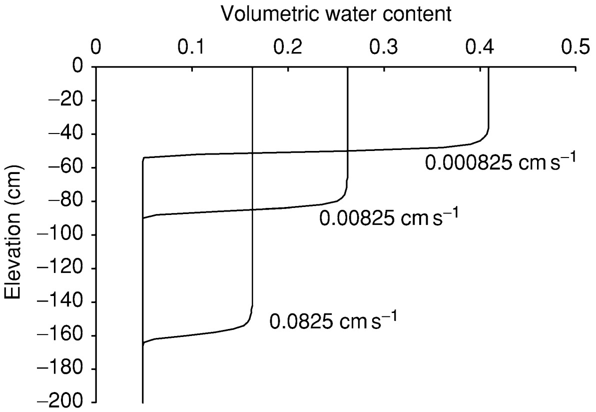


Figure 4 Water-content profiles after 10 h of infiltration at a rate of 0.0005 cm s^{-1} into a sand with a saturated hydraulic conductivity of 0.00086, 0.0086, or 0.086 cm s^{-1} and an initial pressure head of −100 cm.

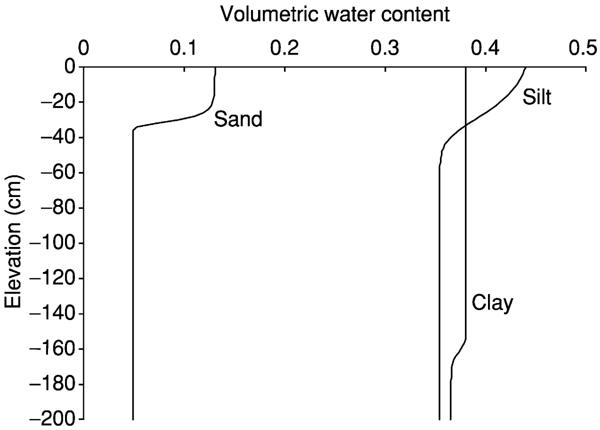


Figure 5 Water-content profiles after 44.4 h of infiltration at a rate of 0.000015 cm s^{-1} into a sand, silt, or clay at an initial pressure head of −100 cm.

The Effects of Gravity

The preceding discussions have considered only vertical infiltration. However, water can also move horizontally across vertical soil surfaces (e.g., the walls of a water-filled surface depression). All of the influences described above for vertical flow have the same effect on horizontal advance in response to water applications controlled by the rate of application at a soil surface. Definite differences arise when water is applied at the soil surface at a constant pressure head (type I boundary condition). In that case, horizontal flow responds only to the pressure gradients caused by the surface-water pressure head, while vertical flow responds to this gradient and to gravity. After a short time, when the pressure-head gradient is high across the surface, the rates of horizontal and vertical flow are nearly equal. However, later the pressure-head gradient approaches zero while the gravitational gradient is always 1. As a result, horizontal flow decreases toward zero while vertical flow reaches a steady nonzero rate. This difference in the horizontal and vertical flow in response to the same applied pressure-head is critical to an understanding of multidimensional infiltration.

Two-Dimensional Infiltration

The preceding discussions have focused on one-dimensional, vertical infiltration. While this simplified representation is applicable to many common conditions (e.g., rainfall or sprinkler irrigation over a flat field), other conditions require a consideration of multidimensional flow. Multidimensional flow can be understood as a balance between horizontal and vertical flow. For instance, a wetting front will proceed farther downward if flow is one-dimensional vertical, than if flow is multidimensional, because in the latter case the amount of water moving vertically downward will be less as a result of water 'lost' to horizontal flow. Figure 6 shows the water-content profiles beneath a point of constant infiltration for both one-dimensional (solid lines) and two-dimensional (dashed lines) flow regimes. Regardless of the soil texture, the wetting front moves deeper under one-dimensional flow conditions. This difference is more pronounced for the finer-textured soils, because their relatively low hydraulic conductivities preclude rapid, downward flow. As a result, the effects of gravity are minimized, and horizontal and vertical flow diverge less slowly. A measure of the effects of horizontal flow on the downward advance of the wetting front can be determined from the total change in water storage directly beneath the point of water application for one- and two-dimensional simulations. For the examples shown in Figure 6, the change in storage directly beneath the point of water application for the sand undergoing two-dimensional flow is 74% of that for the one-dimensional case; the silt and clay show 51% and 32%, respectively. This emphasizes the relatively large impact of horizontal flow on finer-grained, less-permeable media.

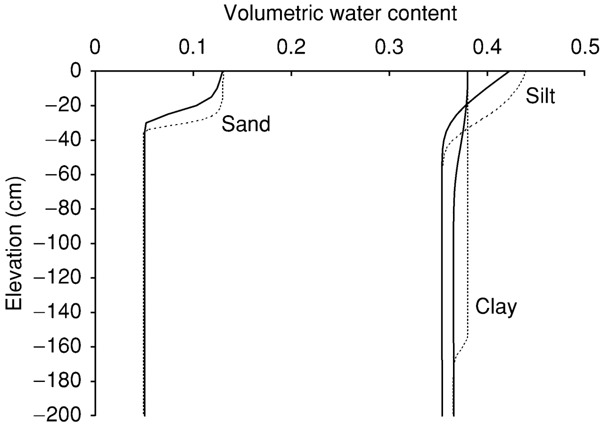


Figure 6 Water-content profiles after 44.4 h of infiltration at a rate of 0.000 015 cm s^{-1} into a sand, silt, or clay at an initial pressure head of −100 cm. Dashed lines show the response calculated for one-dimensional flow. Solid lines show the response directly beneath the point of water application when flow is two-dimensional.

There is a further difference between one-dimensional and multidimensional flow. If flow is one-dimensional, the wetting front continues to advance downward as long as water is applied at the ground surface. Specifically, because the water content behind the wetting front is nearly constant, the flux across the wetting front is equal to the flux across the ground surface. The wetted region can be seen to advance due to this flux across the wetting front, which results in water filling pores that were previously filled with air. In contrast, under multidimensional flow, the wetted region continues to expand in multiple directions. As a result, the surface area of the wetted region continues to increase. The area of the surface bounding the wetted region increases in direct proportion to the flux across the boundary of the wetted region. As a result, at long times, the area of the boundary of the wetted region becomes so large that the flux across the boundary approaches zero at every location and the size of the wetted region becomes essentially constant in time. This gives the unusual result that there can be a constant water content distribution with both depth and time at long times during multidimensional infiltration.

Further Considerations

The preceding discussions have been limited to infiltration into a homogeneous, isotropic porous medium overlying a deep water table. Infiltration in the presence of a shallow water table obeys the same general principles of Darcy's law and the Richards equation. When a shallow water table is present, the moisture profiles (and pressure profiles) must merge vertically with the saturated profile and capillary fringe. More complex aspects of infiltration arise due to the effects of horizontal and subhorizontal layering, capillary barriers, preferential flow, unstable finger-flow, anisotropy, dual porosity, changes in hydraulic properties due to clogging with time, and the effects of air resistance during infiltration.

See also: **Compaction**; **Crusts:** Structural; **Hydrodynamics in Soils**; **Water Cycle**

Further Reading

Green WH and Ampt GA (1911) Studies on soil physics. I. Flow of air and water through soils. *Journal of Agricultural Science* 4: 1–24.

Hillel D (1998) *Environmental Soil Physics*. New York: Academic Press.

Jury WA, Gardner WR, and Gardner WH (1991) *Soil Physics*, 5th edn. New York: John Wiley.

Philip JR (1969) Theory of infiltration. *Advances in Hydroscience* 5: 215–305.

Richards LA (1931) Capillary conduction of liquid through porous media. *Physics* 1: 318–333.

Schaap MG, Leij FJ, and van Genuchten MT (2001) ROSETTA: a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. *Journal of Hydrology* 251: 163–176.

Simunek J, Šejna M, and van Genuchten MT (1999) The HYDRUS-2D software package for simulating two-dimensional movement of water, heat, and multiple solutes in variably saturated media. Version 2.0, IGWMC-TPS-53. International Ground Water Modeling Center. Golden, CO: Colorado School of Mines.

Smith RW (2002) *Infiltration Theory for Hydrologic Applications*. Water Resources Monograph 15. Washington, DC: American Geophysical Union.

Warrick AW (2003) *Soil Water Dynamics*. New York: Oxford University Press.

Infrared Spectroscopy *See* **Fourier Transform Infrared Spectroscopy**

IRON NUTRITION

K Mengel and H Kosegarten, Justus Liebig University, Giessen, Germany

Introduction

Iron (Fe) is an indispensable element for all living organisms, in which it is present in many enzymes. Some of the most important enzymic reactions depend on changes between the divalent and trivalent Fe forms to enable the transfer of electrons as occurs in the respiratory and photosynthetic electron (e^-) transport chains. Fe taken up from the soil by plants is for many animal species the exclusive nutritional source of Fe, and in vegetables and fruits it is also of high importance in the human diet. Apart from organic soils containing virtually no inorganic soil matter, Fe is present in soils in abundant quantities. Its availability to plants, however, depends much on ecologic conditions. In aerobic soils its concentration in the soil solution is extremely low and decreases with an increase in soil pH. Of a still higher importance for Fe availability to plant roots is the soil microbial activity by which siderophores are produced which dissolve mineral Fe. Under anaerobic soil conditions, mineral Fe(III) is reduced and the Fe^{2+} produced is soluble and may attain toxic concentrations under soil flooding. Fe nutrition of crops may be insufficient in calcareous and alkaline soils caused by an inhibition of Fe(III) reduction in the root apoplast. Iron deficiency affects crop growth and quality.

Iron in the Soil and its Acquisition by Plants

Of all the chemical elements composing the Earth's crust, Fe is present fourth-highest by weight, with mean concentration of approximately $50\,\mathrm{mg\,kg^{-1}}$. The greatest part of soil Fe occurs in the crystal lattices of numerous primary soil minerals. In 2:1 phyllosilicates such as mica, smectite, and illite, it is located in the center of the octahedra in divalent and trivalent form, depending on the redox conditions of the soil. Weathering of these primary Fe-containing minerals results in the production of various Fe oxides such as goethite, hematite, and ferrihydrate. These oxides are the most important direct Fe sources for plants and soil microorganisms and their solubility decreases in the following sequence:

$$\begin{aligned}&\mathrm{Fe(OH)_3}\text{ amorphous} > \mathrm{Fe(OH)_3}\\&\quad\text{in soils(semiamorphous)} > \gamma\text{-}\mathrm{Fe_2O_3}\text{ maghemite}\\&\quad > \gamma\text{-FeOOH lepidocrocite} > \alpha\text{-}\mathrm{Fe_2O_3}\text{ hematite}\\&\quad > \alpha\text{-FeOOH goethite}\end{aligned}$$

Goethite is the most abundant Fe mineral in soils. Another Fe oxide of interest is ferrihydrite ($Fe_2O_3 \cdot 9H_2O$). Heterotrophic soil microorganisms are able to oxidize the organic component of Fe chelates, during which process the Fe^{3+} released is hydrolyzed and precipitated:

$$Fe^{3+} + 3H_2O \leftrightarrow Fe^{III}(OH)_3 + 3H^+$$

The resulting ferrihydroxide then gives rise to other Fe oxides, particularly ferrihydrite. Under aerobic soil conditions, the solublity of all Fe(III) oxides in water is extremely low and increases as soil pH decreases. Accordingly a pH as low as 3 is required for inorganic Fe solubility sufficient to cover plant demand.

Since most higher plants that grow on aerobic soils do so over a pH range well above 3 and are usually adequately supplied with iron, it can be concluded that organic, water-soluble Fe compounds (siderophores) must contribute significantly to the Fe supply of plants from soil. Total soluble Fe concentration in soils should be at least 1 μmol l^{-1}. As shown below in more detail, numerous microbial species and also some higher plant species produce organic molecules (siderophores) that chelate the Fe of insoluble inorganic Fe compounds, thus forming soluble Fe compounds (Fe siderophores), which are indispensable for the Fe supply of plants in aerobic soils.

The situation is quite different under anaerobic soil conditions, where soil microorganisms use the Fe(III) of the various Fe(III) oxides as a terminal e^- acceptor in the respiratory chain as a substitute for O_2:

$$Fe^{III}(OH)_3 + e^- \rightarrow Fe^{2+} + 3OH^-$$

The Fe^{2+} is water-soluble and may accumulate to toxic concentrations, especially in flooded soils. Plant toxicity begins at redox potentials of less than 0.5 V.

In aerobic soils, Fe availability to plants depends crucially on the formation of Fe-chelating organic molecules, so-called siderophores. These are produced by soil bacteria, soil fungi, and some plant species. More than 100 different siderophores are known that are produced by soil fungi and soil bacteria. Numerous siderophores are derived from hydroxamate, which binds Fe(III) with a coordinative and an ionic bond. These siderophores are water-soluble and stable over a wide pH range and thus also stable in the alkaline pH range characteristic of calcareous and alkaline soils, where the solubility of Fe oxides is extremely low. Grasses are able to synthesize siderophores in their roots (phytosiderophores). Phytosiderophores such as mugineic acid and its derivatives avenic acid, 3-hydroxymugineic acid, and distichonic acid are nonprotein amino acids that are capable of binding ferric Fe. As shown in Figure 1, they have six functional groups by which they bind Fe(III) in a hexadentate form, with three carboxyl groups, two N atoms, and one OH. Phytosiderophores are secreted by root tips a few millimeters behind the apex. Secretion of phytosiderophores depends on the Fe nutritional status of plants and is particularly high if plants are suffering from an insufficient Fe supply.

Phytosiderophores thus released by roots react with the Fe(III) present in soil minerals to form an Fe(III) siderophore complex, which is soluble in the soil solution and can be translocated by mass flow or diffusion toward the plant roots. There is no firm evidence yet that the Fe(III) siderophore complex is translocated across the plasmalemma of the root cortical cells into

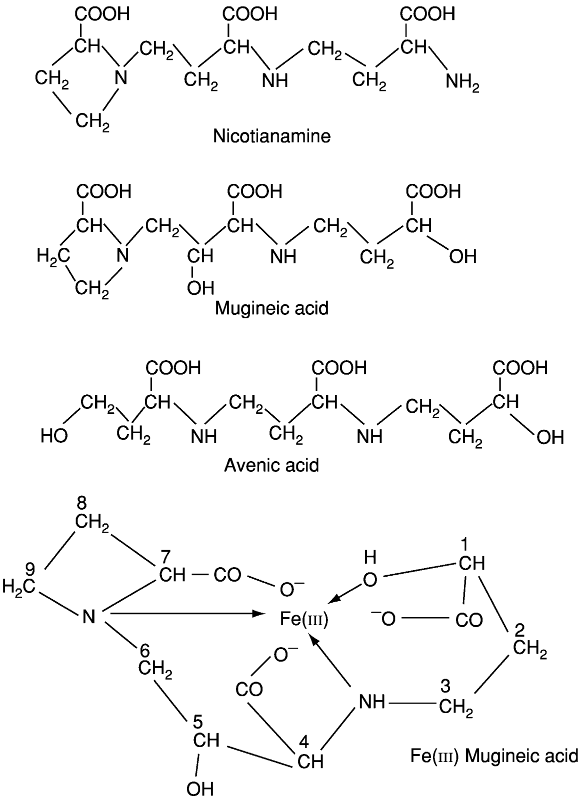


Figure 1 Important siderophores and the hexadentate Fe(III) mugineic acid complex.

the cytosol, because until now neither the transporter nor the genes coding for such a system have been found. Due to their capability for producing siderophores, grasses are better equipped for mining soils for Fe than are dicotyledenous species. Nicotianamine, shown in Figure 1, chelates Fe^{2+}. This complex is of no importance for the mobility of Fe in soils but is essential for the transport of Fe in the cytosol.

One may distinguish between two strategies by which plant species are able to cope with an insufficient Fe supply. In strategy-I plants, the dicotyledons respond by increased H^+ secretion from the roots, thus mobilizing soil Fe and depressing apoplastic pH, promoting Fe uptake. The grasses (strategy-II species) respond by enhanced release of phytosiderophores from the roots. Monocotyledonous species differ in their potential to release phytosiderophores and similarly there may be differences between cultivars and genotypes in their abilities to synthesize and release siderophores. Siderophores including phytosiderophores are decomposed by soil microorganisms. There is direct evidence that microbial activity in soils is indispensable for the acquisition of Fe by dicotyledonous species and is also of some importance for grasses.

Both dicotyledons and grasses form rhizodermal transfer cells under the conditions of insufficient Fe supply. These cells are characterized by invaginations and thus by an enlargement of the plasma membrane, and therefore by an increase in Fe-uptake systems, including H^+ pumps. The apoplastic pH is of crucial importance for Fe uptake. The increased H^+ secretion by roots under conditions of insufficient Fe promotes Fe uptake as a consequence of the decrease in the apoplastic pH. Proton secretion by roots may also promote the solubilization of inorganic soil Fe. In most cases, however, H^+ ions diffusing from the roots into the rhizosphere are quickly neutralized. This is particularly true for calcareous and alkaline soils high in pH.

There exist various methods for the determination of Fe bioavailability. One of the most used techniques is that of soil extraction by diethylenetriaminepentaacetic acid (DTPA). This extractant reacts with soluble and exchangeable Fe, as well as with inorganic insoluble Fe such as Fe oxides. The problem with all Fe extraction tests is that the bioavailability of soil Fe depends more on the presence and activity of siderophores than on the physicochemical solubility and extractability of soil Fe.

Iron Uptake, Long-Distance Transport in the Plant, and Storage in the Cell

Iron has a unique uptake system, as the transport across the plasma membrane is associated with the reduction of Fe(III). Uptake takes place in the roots and in the mesophyll. It is supposed that the principal uptake mechanism of both is the same; the roots, however, are supplied by a wide range of different siderophores with chelated Fe(III). In the leaves it is mainly Fe(III) dicitrate which is translocated via the xylem, through the larger and minor vein system, and the free space of cell walls (apoplast) into the leaf tissues, where the Fe is taken up into the cells. The term 'uptake' means primarily the transport of Fe across the plasma membrane. Fe(III) reductase and the Fe^{2+} channel are essential components of the uptake system, as shown in Figure 2. Fe(III) reductase is a transmembrane enzyme, contacting with one part the cytosol and with the other the apoplast. The cytosol supplies the reductase with reducing power in the form of reduced nicotinamide adenine dinucleotide phosphate (NAD(P)H). The reducing enzyme activity is highly sensitive to the apoplastic pH and decreases as the pH increases, as shown by direct apoplastic pH measurements in sunflower leaves. Generally the optimum apoplastic pH range is provided by the activity of the plasmalemma proton pump (Figure 2). Processes that change the apoplastic pH, such as the H^+–nitrate cotransport across the plasmalemma, raise the apoplastic pH and thus depress Fe(III) reduction and Fe^{2+} uptake. Highest reduction rates were found at $pH < 5.44$. Basically the same relationship between pH in the apoplast and the Fe(III) reduction is supposed to exist in roots. Here it is particularly HCO_3^- that increases the apoplastic pH and thus restricts or even blocks Fe(III) reduction. Monohydrogen carbonate (HCO_3^-) is the anion species dominant in soils high in pH, such as calcareous and alkaline soils, and these are the soils that are prone to Fe deficiency. As shown in Figure 2, the apoplastic pH is dependent on the H^+ pump activity. Accordingly cultivars resistant to Fe chlorosis release more H^+ from the roots than do the susceptible ones.

Iron uptake of roots is restricted to the root tips, and the Fe uptake systems are located in the plasma membranes of the epidermal cell layer. From this

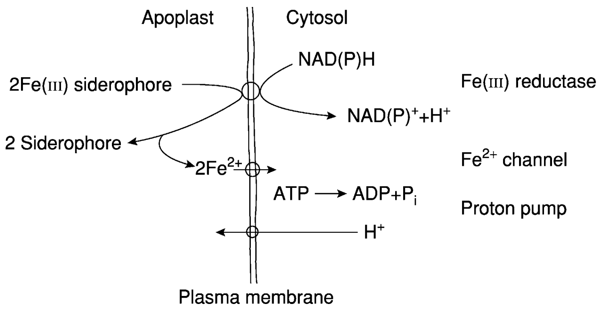


Figure 2 Scheme of Fe^{2+} uptake across the plasmalemma.

follows that the surface area from which Fe is taken up is small in comparison with the total surface area exposed to the soil by young roots from which most other plant nutrients are absorbed. There are indications that in leaves the Fe uptake systems are also exclusively located in the meristematic tissues. The Fe^{2+} produced in the apoplast passes through a specific channel of the plasma membrane into the cytosol. This takes place by diffusion and thus depends on the electrochemical gradient for Fe^{2+} across the plasmalemma. In flooded soils, where a surplus of Fe^{2+} may be produced by microbial respiratory activity, relatively high rates of Fe^{2+} may diffuse into the cytosol and thereby induce Fe toxicity. Toxicity may arise, because the excess Fe^{2+} easily reacts with O_2, forming the highly toxic superoxide radical O_2^-:

$$Fe^{2+} + O_2 \rightarrow O_2^- + Fe^{3+}$$

Under normal conditions the cytosolic Fe^{2+} concentration is maintained at a low level by the translocation of Fe^{2+} into the plastids, where it is oxidized by O_2, as shown in Figure 3. The resulting superoxide radical is detoxified by superoxide dismutase. The Fe^{3+} produced in the oxidation process reacts with water and thus forms the Fe(III) hydroxide:

$$Fe^{3+} + 3H_2O \rightarrow Fe^{III}(OH)_3 + 3H^+$$

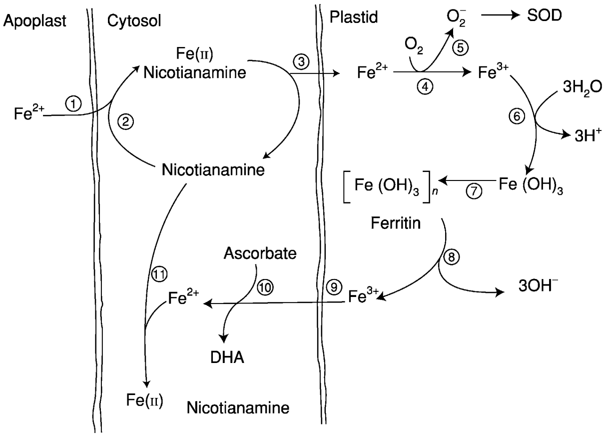


Figure 3 Putative scheme of Fe uptake into the cytosol and into the plastid, Fe incorporation into ferritin, and Fe release from ferritin into the cytosol:

1. Fe uptake across the plasmalemma;
2. Fe binding by nicotianamine;
3. Fe^{2+} transport across the plastid membrane;
4. Fe^{2+} oxidation to Fe^{3+};
5. Formation of the superoxide radical;
6. Formation of $Fe(OH)_3$;
7. Incorporation of $Fe(OH)_3$ into ferritin;
8. Release of Fe^{3+} by ferritin;
9. Transport of Fe^{3+} across the plastid membrane;
10. Reduction of Fe^{3+} by ascorbate to dehydroxyascorbate (DHA);
11. Binding of Fe^{2+} by nicotianamine.

The Fe(III) hydroxide produced is the basic unit for the formation of ferritin, which is an important Fe storage form in plants, animals, and microorganisms. The core of ferritin consists mainly of Fe(III) hydroxide and is surrounded by a protein layer, apoferritin. The core comprises several thousand Fe(III) atoms. This Fe store is characterized by the storage and the release of Fe according to physiological demand and thus represents an Fe buffer. In higher plants, chloroplasts may store remarkably high quantities of Fe so that their Fe concentrations may be 5–10 times higher than those in the corresponding leaf tissue.

Under conditions of physiological Fe demand, Fe^{3+} diffusing out of the plastid is reduced by ascorbate, and the resulting Fe^{2+} forms an Fe(II) chelate with nicotianamine (Figure 3).

Nicotianamine is a nonprotein amino acid which is structurally related to siderophores (Figure 1) and forms stable complexes with Fe^{2+}. Fe–nicotianamine is not involved in membrane transport and therefore cannot be excreted into the xylem. Fe^{2+} nicotianamine is essential for the symplastic transport from cell to cell. Fe–nicotianamine is phloem-mobile. Until now, however, it has not been known whether symplastic phloem-loading of Fe plays a major role, since older leaves hardly provide sufficient Fe to younger leaves. Indeed it is for this reason that Fe-deficiency symptoms first appear in younger leaves.

Biochemical Functions of Fe in Plants

In most of the Fe-containing enzymes, the functional mechanism is based on the valency change of Fe. The enzyme with the highest negative standard redox potential is ferredoxin ($E'_o = -0.43\,V$). Ferredoxins have a low molecular weight of approximately 6–12 kDa. They are stable 2Fe–2S proteins, with a prosthetic group that contains two Fe and two inorganic S atoms, which bridge the two Fe atoms. Fe is further bound by two bonds to cysteine residues. A soluble form of ferredoxin exists as well as a form integrated in the thyalkoid membrane, which is also known as the 2Fe–2S complex. The redox reaction is a one-e^- step process, which means that in the reaction only one and not two e^- are released or taken up during electron transport. In the oxidized form, both Fe constituents are present as Fe(III), and in the reduced form only one Fe is present as Fe(II).

During photosynthesis the soluble ferredoxin present in the chloroplast receives the e^- directly from the membrane-bound ferredoxin of photosystem I and transfers it to other important enzymes related to nutrient assimilation, such as $NADP^+$ (CO_2 assimilation), nitrite reductase (N assimilation), sulfate reductase (S assimilation), heme proteins (cyclic photophosphorylation), as well as to O_2. Reduction of O_2 by ferredoxin takes place under high light intensity associated with a surplus of photosynthetic e^-. In this case the superoxide radical is formed, which, if not detoxified completely by superoxide dismutase, may lead to the damage of photosynthetic pigments, known as photoxidation. Ferredoxin is also essential for the N_2 assimilation of free-living and symbiotically living N_2-fixing bacteria. The ferredoxin provides the e^- for the N_2 reduction (nitrogenase). Besides the binuclear Fe–S complexes, tetranuclear Fe–S complexes (4Fe–4S) also exist that catalyze redox reactions in an analogous way. The core of nitrogenase consists of an Fe4–S4 cluster and an Fe3–Mo–S4 cluster. In the latter, one Fe is substituted by one Mo.

The enzymes discussed above belong to the nonheme Fe group. The group of heme Fe enzymes is characterized by heme as the prosthetic group with Fe in the center of the tetrapyrrole ring, as occurs in the cytochromes which mediate e^- transport in the photosynthetic and respiratory e^- transport chains. Other heme enzymes include catalase and peroxidase.

Some heme enzymes do not react by a valency change but function as O_2 transporters such as myoglobin and hemoglobin in animal organisms and leghemoglobin in the symbiotic nitrogen fixation of N_2-fixing bacteria and their host plants. The leghemoglobin brings about a controlled delivery of O_2 to the N_2-fixing bacteroid. This is essential because the nitrogenase is very sensitive to O_2.

The pathway for the synthesis of chlorophyll requires Fe, therefore insufficient Fe supply frequently results in chlorosis, which means a lack of chlorophyll. Ribonucleotide reductase is another Fe-containing enzyme of universal importance, because it catalyzes the reduction of ribonucleotide to deoxyribonucleotide. These are the building blocks for DNA. Insufficient DNA affects cell division and thus growth.

Lime-Induced Chlorosis

Absolute Fe deficiency in plants grown in the field is very rare. However, it may occur on organic soils with a very low proportion of mineral soil matter. Lime-induced chlorosis occurs almost exclusively on calcareous and alkaline soils. It has been recognized since the nineteenth century and may cause considerable loss in crop production. Deficiency symptoms are initially the yellowing of intercostal areas in the youngest leaves, which proceeds until the total leaf is almost golden-yellow except for small green stripes along the ribs. In extreme cases the whole foliage of a plant may become yellow (chlorotic).

Calcareous and alkaline soils have a high potential to produce HCO_3^-, which may diffuse into the root apoplast and neutralize protons released by the plasmalemma H^+ pump, thus increasing the apoplastic pH, a process associated with a decrease in Fe(III) reductase activity and Fe^{2+} uptake. This interpretation is in agreement with earlier findings in which 24 soils were investigated differing in carbonate concentration in relation to chlorosis of sorghum. Of the various parameters tested, the intensity of visual chlorosis in sorghum leaves was most closely correlated ($P<0.001$) with soil pH in the KCl extract and with the alkalinity of soils measured as the concentration of HCO_3^-, which was in the range of 1.0–4.5 mol l^{-1} in the soil extract. Iron uptake of plants was negatively correlated ($P < 0.001$) with the pH in the KCl extract and with the HCO_3^- concentration in soils. The latter also depends on the water saturation of soils, high saturation increasing the partial pressure of CO_2 in soils and thus the dissolution of CO_2 and the production of HCO_3^-. Under such highly saturating conditions, chlorosis may occur and is known as 'bad-weather chlorosis.'

Frequently, chlorotic plants have high Fe concentrations in their roots. It has been suggested that this high Fe concentration is due to adhering soil particles on the root surface. However, plants grown in solution culture where no adhering soil particles are present may also have Fe concentrations in the roots ranging from 300 to 600 mg kg^{-1} dry weight, which is several times higher than the Fe concentration in leaves. This observation shows that plants are capable of storing high amounts of Fe in their roots. Such high Fe concentrations in roots occur when the Fe translocation from roots to shoots is restricted. By means of EDX analysis, it has been shown that the Fe concentration in the radial cell walls of the first cell layer of the root cortex parenchyma of maize grown in a calcareous soil (Mollisol) is more than 10 times higher than that of plants grown in an acid organic soil (Histosol). In the radial cell wall of the third cell layer, Fe concentrations are lower and there are no major differences between the two soils. Interestingly, the DTPA-extractable Fe concentration in the Histosol amounted to 499 and in the Mollisol to 4.2 mg Fe kg^{-1} soil. Nevertheless, Fe accumulates in the root apoplast of the plants grown on the calcareous soil. Since the Fe is located in the radial cell wall, this Fe accumulation cannot be caused by soil contamination. This phenomenon of Fe accumulation is of particular interest, because selective Fe^{2+} channels are exclusively located in the plasmalemma of the root epidermis. If the root apoplastic pH is high because of abundant HCO_3^-, $Fe^{(III)}$ reduction and Fe uptake are restricted. Finally Fe siderophores are decomposed by soil microorganisms and the resulting Fe^{3+} is bound to cell walls. Toxic Al concentrations may also cause extremely high Fe concentrations in roots, whereas the translocation of Fe into the shoot is restricted. Since toxic cationic Al species inhibit the plasmalemma H^+ pump, the resulting high apoplastic pH presumably depresses Fe(III) reduction and hence Fe^{2+} uptake.

A further characteristic of Fe chlorosis is that the intensity of the chlorosis, i.e., the chlorophyll concentration, is not correlated with the Fe concentration in leaves. Leaves with the same Fe concentration may show chlorosis whereas others of the same species may not. Based on pH measurements and Fe(III) reduction in leaves of sunflowers, a pH >5.5 in the leaf apoplast depresses Fe(III) reduction and thus Fe uptake into the cytosol. Hence in chlorotic leaves with a relatively high Fe concentration a substantial part of this Fe may be located in the leaf apoplast and therefore is of no physiological use. Nitrate nutrition promoting leaf chlorosis increases the apoplastic pH in leaves because of H^+–nitrate cotransport. Accordingly, an acid foliar spray has led to a regreening of chlorotic leaves without increasing the leaf Fe concentration. In parallel investigations it has been shown that citric acid application decreases the apoplastic pH by approximately 0.5 pH units. This finding is in accord with earlier investigations in which a different *in situ* pattern was found for Fe between chlorotic and green leaves by Mössbauer spectrometry. In chlorotic leaves obtained with nitrate nutrition, Fe was present exclusively as Fe(III), whereas, in green leaves obtained with ammonium nutrition, more than 20% of the Fe occurred as Fe(II) and the rest was Fe(III).

In solution culture experiments with sunflower, Fe chlorosis is found when plants are exclusively supplied with nitrate. In the presence of ammonium in the nutrient solution, however, no chlorosis appears. Since ammonium is taken up preferentially relative to nitrate, nitrate may be translocated to the leaves only at minute rates, so that a major pH increase in the leaf apoplast due to H^+–nitrate cotransport will not appear. The phenomenon of leaf chlorosis being associated with high Fe concentration in leaves is only found on soils with a high pH. On such soils ammonium does not accumulate in the soil, because the equilibrium $NH_3 + H^+ \leftrightarrow NH_4^+$ is shifted to the ammonia side of the equation and ammonia is released from the soil by evaporation. Hence, under such conditions, plants may be exclusively supplied with nitrate, which may lead to a high apoplastic leaf pH, restricting Fe(III) reduction and leading to an accumulation of Fe in the leaf apoplast. Such conditions may especially appear under dry weather, which favors the volatilization of NH_3 and CO_2 from the soil. The CO_2 release leads to a decrease in the soil HCO_3^- concentration and hence benefits the Fe^{2+} uptake by roots; the lack of NH_4^+ promotes nitrate uptake of the roots and thus the nitrate transport to the leaves, with its adverse effect on Fe uptake from the leaf apoplast into the cytosol. Under such conditions Fe may accumulate in the leaf apoplast the longer such conditions prevail. Therefore high leaf Fe concentrations in chlorotic leaves appear mainly in older leaves and later in the season. Iron accumulation is enforced by a depression in leaf growth because of a lack in Fe in the cytosol. Iron deficiency on calcareous soils not only is evident in leaf chlorosis, but may also result in a serious growth reduction without visible symptoms of chlorosis.

In natural vegetation there is hardly any evidence of plant species with lime-induced chlorosis, since genotypes have been developed during evolution which have adapted to the environmental conditions, as shown for *Lathyrus* species. *Lathyrus montanus* is susceptible to Fe chlorosis, whereas *L. pratense*, evolved on calcareous soils, is not. This is also true for *Vitis* species, of which rootstocks are used in wine production. Rootstocks of *Vitis labrusca* and *V. riparia*, both evolved in the north and northeast of the USA, are susceptible to chlorosis; but *V. berlandieri*, which evolved on calcareous soils in Texas and Mexico, is resistant. The physiological processes relating to the chlorosis resistance are not yet completely understood. The rootstock of the resistant vine cultivar Faber releases more than twice as much H^+ by its roots than the chlorosis-susceptible cultivar Huxel. This H^+ release reflects the apoplastic pH in roots and therefore also the conditions for Fe(III) reduction as considered above. Even wheat cultivars, which supposedly acquire Fe mainly by phytosiderophores, differ considerably in their resistance to chlorosis. When sunflowers and *Vicia faba* were grown under the same conditions in solution-culture with nitrate $+HCO_3^-$ the sunflowers became chlorotic, but *V. faba* did not. This behavior is related to the effect of nitrate $+HCO_3^-$ on the pH of the xylem sap, which was raised in sunflowers but not in *Vicia*. Supposedly *Vicia* must pump relatively high rates of H^+ into the xylem sap so that the pH remains low even in the presence of nitrate $+HCO_3^-$ supply.

Iron Deficiency and Toxicity

Growing dicotyledon species in a sterile soil results in growth-retarded plants with low Fe concentrations in the leaves. With sunflowers, typical leaf chlorosis appears. Rape plants do not develop chlorosis, but growth is depressed and the leaves remain very small. The addition of an Fe chelate to the soil of the sterile grown plants produces vigorous growth, proving that the growth restriction is attributable to Fe deficiency. Also, vines may suffer from an inadequate supply of Fe without showing chlorotic symptoms. Chlorosis is a very spectacular symptom of Fe deficiency, but with only symptoms of reduction in growth, in plants of normal habit and green leaves, it is hardly possible to recognize a lack of Fe as a cause of deficiency. Supposedly the poor growth is caused by an insuffficent synthesis of deoxyribonucleotide and thus restricts deoxyribonucleic acid (DNA) synthesis.

Plants deficient in Fe have reduced activities of catalase, superoxide dismutase, and ascorbic acid oxidase. Iron deficiency causes a disturbance of thyalkoid formation in chloroplasts and thus affects photosynthetic activity. Iron toxicity in plants may occur on flooded soils because of a high Fe^{2+} concentration so that Fe^{2+} diffuses uncontrolled into the cytosol. Symptoms of Fe toxicity in rice are known as 'bronzing,' which in the beginning is characterized by tiny brown spots on the leaves which later develop into a uniform brown color. The Fe concentration in such leaves is high and in the range of 300–1000 mg Fe kg^{-1} dry matter. Iron toxicity in rice is known in various rice-growing areas and occurs especially on heavy soils.

Iron Application

Curing Fe chlorosis of plants grown on calcareous or alkaline soils may present difficulties, since chlorosis in most cases is a physiological deficiency and roots or even leaves may be rich in Fe but still suffering from insufficient Fe in the cytoplasm. Generally Fe chelates are supplied as a soil fertilizer or as a foliar spray. The most important Fe chelate in this respect is EDDHA (ethylenediaminedi(*o*-hydroxyphenylacetic)acid), which is stable over a pH range of 4–10. For tree crops a foliar spray may cure the deficiency.

Soil applications of mixtures of Fe salts with organic matter such as manure, compost, peat, and sewage sludge may help to cure Fe chlorosis of crops. From the salts, vivianite ($\text{Fe}_3^{\text{II}}(\text{PO}_4)_2$) is of particular interest, as its bivalent Fe can be directly taken up by roots. The problem, however, is that in aerobic soils Fe^{2+} is quickly oxidized to Fe^{3+} and as such reacts immediately with water to Fe(III)(OH)_3, which precipitates.

List of Technical Nomenclature

E'_0	Standard redox potential (volts)

Further Reading

Abadia J (ed.) (1995) *Iron Nutrition in Soils and Plants*. Dordrecht, the Netherlands: Kluwer Academic.

Becker R, Grün M, and Scholz G (1992) Nicotianamine and the distribution of iron in the apoplasm and symplasm of tomato (*Lycopersicon esculentum* Mill.). *Planta* 187: 48–52.

Chen Y and Barak P (1982) Iron nutrition of plants in calcareous soils. *Advances in Agronomy* 25: 217–240.

Chen Y and Hadar Y (eds) (1991) *Iron Nutrition and Interactions in Plants*. Dordrecht, the Netherlands: Kluwer Academic.

Crowley DE, Wang YC, Reid CPP, and Szaniszlo PJ (1991) Mechanism of iron acqusition from siderophores by microorganisms and plants. In: Chen Y and Hadar Y (eds) *Iron Nutrition and Interactions in Plants*, pp. 213–232. Dordrecht, the Netherlands: Kluwer Academic.

Kosegarten H and Koyro HW (2001) Apoplastic accumulation of iron in the epidermis of maize (*Zea mays*) roots grown in calcareous soil. *Physiologia Plantarum* 113: 515–522.

Kosegarten H, Hoffmann B, and Mengel K (1999) Apoplastic pH and Fe^{3+} reduction in intact sunflower leaves. *Plant Physiology* 121: 1069–1079.

Lindsay W (1991) Iron oxide solubilization by organic matter and its effect on iron availability. In: Chen Y and Hadar Y (eds) *Iron Nutrition and Interactions in Plants*, pp. 29–36. Dordrecht, the Netherlands: Kluwer Academic.

Loeppert RH and Hallmark CT (1985) Indigenous soil properties influencing the availability of iron in calcareous soils. *Soil Science Society of America Journal* 597–603.

Loeppert RH and Inskeep WP (1996) Iron. In: Bigham JM (ed.) *Methods of Soil Analysis, part 3, Chemical Methods*, pp. 639–664. Madison, WI: Soil Science Society of America.

Machold O, Meisel W, and Schnorr H (1968) Bestimmung der Bindungsform des Eisens in Blättern durch Mössbauer-Spektrometrie. *Naturwissenschaften* 10: 499–500.

Marschner H, Treeby M, and Römheld V (1989) Role of root induced changes in the rhizosphere, for iron acquisition in higher plants. *Zeitschrift für Pflanzenernährung und Bodenkunde* 152: 197–204.

Mengel K (1994) Iron availability in plant tissues – iron chlorosis on calcareous soils. *Plant and Soil* 165: 275–283.

Neilands JB and Leong SA (1986) Siderophores in relation to plant growth and disease. *Annual Review of Plant Physiology* 37: 187–208.

Römheld V and Marschner H (1986) Mobilization of iron in the rhizosphere of different plant species. *Advances in Plant Nutrition* 2: 155–204.

Rroco E, Kosegarten H, Harzay F, Imani J, and Mengel K (2003) The importance of soil microbial activity for the supply of iron to millet and rape. *European Journal of Agronomy* 19: 487–493.

Schwertmann U and Taylor RM (1989) Iron oxides. In: Dixon JB and Weed SB (eds) *Minerals in Soil Environments*, pp. 379–438. Madison, WI: Soil Science Society of America.

Tagliavini M and Rombola AD (2001) Iron deficiency and chlorosis in orchard and vineyard ecosystems. *European Journal of Agronomy* 15: 71–92.

IRRIGATION

Contents

Environmental Effects

S Topcu and C Kirda, University of Cukurova, Adana, Turkey

Introduction

The delta plains of the Euphrates and the Tigris rivers in Mesopotamia, and the Nile in Egypt have witnessed numerous civilizations. With the early practice of irrigated agriculture, these soils have offered a prosperous life to the people of the region. Archeological remains of irrigation structures such as barrages, channels, and deep wells date back to as early as 4000–6000 BC. The real advancement of irrigated agriculture, however, started in the twentieth century with the development of large irrigation projects in Egypt, Pakistan, India, and the USA. Irrigation contributed significantly to the high crop yields achieved following the Green Revolution in the 1960s, which is commonly attributed to breeding of high-yielding crop cultivars and use of fertilizers and pesticides.

The demand for fresh water increased fourfold from 1950 to 1960 with the implementation of new irrigation projects and the increase in municipal and industrial water demands. Irrigated agriculture, complemented by fertilizer and pesticide use, has increased crop yields. In some instances, however, this has been at the risk of soil degradation and water pollution and loss.

Irrigation development projects facilitate improvement of infrastructure such as roads, telecommunications, and energy networks; improvement of public services in health care and education; promotion of investments in agribusiness and related areas; and creation of new employment opportunities. While the country-wide prices of agricultural produce tend to decrease under irrigated agriculture, the living standards of the whole population tend to be improved. Although only approximately 20% of the world's cultivable lands are under irrigation, they account for 40% of global agricultural production. It is projected that the population will reach 8 billion by the year 2025, necessitating a 40% increase in food production. Therefore, it is likely that irrigated agriculture will continue to expand.

The expansion of irrigation projects, however, without consideration of drainage, coupled with insufficient funds for proper implementation of technical and socioeconomic aspects, brings about the question of sustainability. Lack of funds for extension service, lack of training and experience, poor choice of irrigation methods, and excessive use of fertilizers and pesticides all contribute to the unfavorable impacts that irrigation can have not only on soil and water resources, but on the whole environment.

Soon after the transition to irrigated agriculture, the initially achieved increase in crop yields may start to decline, plant diseases that are initially nonexistent may become widespread, and the water table may begin to rise, leading to waterlogging and salinity at a later stage. History has witnessed the disappearance of numerous civilizations not only because of wars and long spells of drought or floods, but also as a result of the unintended consequences of irrigated agriculture. Because of the increased soil erosion from irrigated lands, sedimentation causes filling of irrigation channels and dams. Excess use of water triggers drainage and salinity problems, which may cause irreversible loss of soil fertility. Erosion and sedimentation, as well as salination, have gone hand-in-hand with the deterioration of farmlands and the eventual collapse of some social structures. The Babylonian and Sumerian empires in Mesopotamia are examples

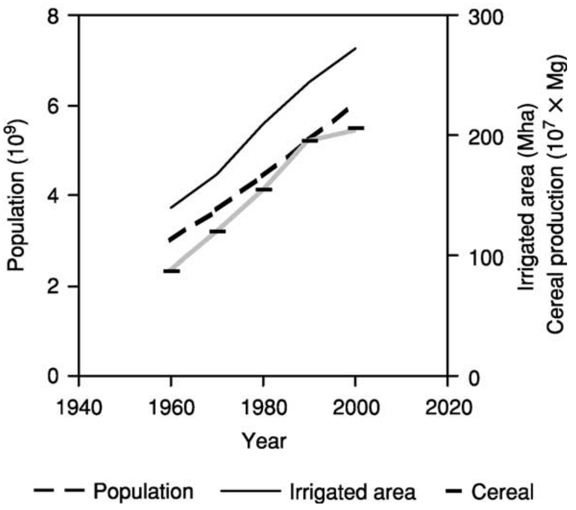


Figure 1 Population, irrigated areas, and annual cereal production (1960–2000). (Adapted from FAOSTAT–Agricultural Data, FAO on-line database.)

of past civilizations whose collapse, in part, was caused by poorly practiced irrigation.

The replenishment of water in the plant root zone during periods when natural rainfall is not sufficient for good plant growth is the classic definition of irrigation. Irrigation is essential to ensure high crop yields in arid and semiarid regions. Supplementary irrigation in humid and subhumid regions not only guarantees high crop yields, but also improves the quality of crops. It is well documented that crop yields under irrigation increase from 100 to 500% compared with yields obtained under rain-fed conditions in arid regions. Starting with the 1950s, irrigated areas increased from a mere 94 Mha to 271.7 Mha in the year 2000. Figure 1 shows the expansion of irrigated areas in the last 50 years and compares it with the global increases in population and cereal production. Because governments had committed to investing in large irrigation projects, the rate of expansion of irrigated areas was 2% annually until the late 1970s. Starting from the early 1980s, it leveled off to around 1% per year in the 1990s. Although irrigated agriculture is practiced in nearly all continents, 70% of the globally irrigated areas exist in Pakistan, China, India, and other developing countries in arid and semiarid regions.

The expansion of irrigated areas has brought about a significant increase in the production of food crops, particularly cereals. Cereal production shows a strong association with the increasing acreage of irrigated areas (Figure 1). Although 31% of the world's cereal-planted area is under irrigation, 42% of global cereal production originates from these areas. Irrigated areas in South and Southeast Asia, and in Near-East Asian countries produce, respectively, 50%, 65%, and 75% of the total cereal production in the respective regions. Irrigated agriculture and its continuing expansion are essential in meeting the needs of increasing populations in the developing counties.

In the future, the largest expansion of irrigated areas will take place in East and Near-East Asian countries and in North Africa where farmlands are limited and irrigation is the only means of increasing agricultural production. However, it is expected that the increase in irrigated areas will be only marginal in Sub-Saharan Africa and Latin America. It has been estimated that three-quarters of the farmlands in East and Near-East Asian countries and North Africa will have irrigation projects implemented in the near future. Similarly, irrigated agriculture will extend to more than 90% of the cultivable lands in South Asia. Thus, it is expected that irrigated areas will reach 330 Mha by the year 2025. The major constraint to the expansion of irrigated areas is the increased demand for municipal and industrial water allocations in competition with agriculture.

Although the benefits of irrigated agriculture are clear, there may be major environmental drawbacks if proper irrigation management is not implemented. Mismanagement of water can cause degradation of soils and pollution of water resources.

Environmental Impact of Irrigated Agriculture

The natural hydrology of watersheds is disturbed during storage and conveyance of water allocated for irrigation, and during discharge of drainage and returning waters from irrigated areas. Overexploitation of groundwater, changing flow regimes of rivers, and raising of water tables (which may trigger salt accumulation) are some of the environmental issues to be addressed. Use of high amounts of pesticides and fertilizers may be a source of pollution of water resources. Furthermore, large bodies of water stored in dams and diversion channels can cause significant changes in local climate.

Water Shortage

Irrigated agriculture uses the highest portion of the world's water resources. The agricultural sector accounted for 66% of the total water withdrawn from both surface water and groundwater resources in 1995 (Figure 2). However, water withdrawal and consumption show great variability in different continents and countries. For example, total water withdrawal in Asia is 57%, of which 80% is used in

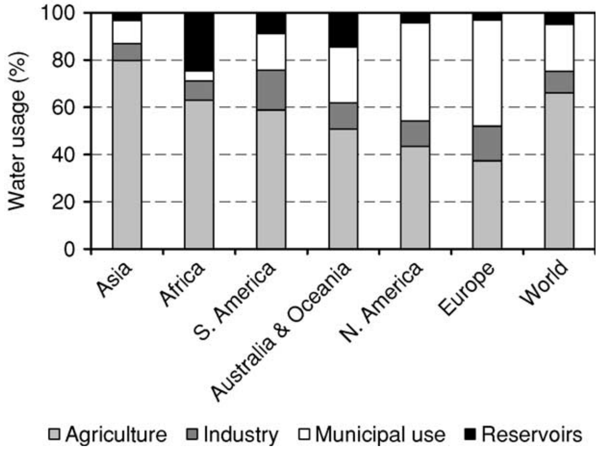


Figure 2 Percentage water consumption of annually withdrawn water resources by sector and continent (1995). (Adapted from Shiklomanov LA (1999) *World Water Resources at the Beginning of the 21st Century*. IHP Report. Paris, France: UNESCO.)

irrigated agriculture. While only 1% of renewable water resources is used in Latin America, the consumption exceeds the renewable limit in Libya and Saudi Arabia.

Actual plant water consumption is only a small portion of the total water diverted to irrigated agriculture because of losses in conveyence networks, use of irrigation methods of low water-application efficiency, excess water application due to unrealistically low water-pricing, and the like. A survey of irrigation networks carried out in 90 different developing countries in 1998 showed that water-use efficiency was as low as 38%. Water allocation for agriculture decreased from 78% to 66% during the last 50 years, in spite of the increasing demand of the agricultural sector. Because of overexploitation of groundwater resources for irrigation in many countries, water table depths have increased. For example, the water table depth in Northern China increased from 8 m to 50 m from 1967 to 2000. Similarly, it was reported that water levels of 10 aquifers in Mexico decreased annually 1.8–3.3 m. More than one million deep wells are opened annually in the Indian peninsula and Pakistan. The total water withdrawal from these wells exceeds the annual recharge rate. Overexploitation of deep wells close to coastal areas triggers seawater intrusion and thereby causes degradation of water quality, which in turn endangers the sustainability of irrigated agriculture.

In the coming decades, three-quarters of the world's population will face water scarcity. Therefore, innovations are needed to sustain high crop yields with lesser amounts of water allocated for irrigated agriculture. There are so-called deficit irrigation practices to sustain high crop yields when irrigation water supplies are scarce and limited. Research for developing new innovative techniques to decrease irrigation water requirements is still continuing. To this effect, a technique called the partial root drying practice, where only half of the plant root zone is wetted by applying water to one side of the rows while keeping the other half dry, is an example of newly emerging innovative irrigation practices to save water. The wetted half of the roots is alternately changed in the subsequent irrigations. The technique is easily practiced in furrow-irrigated crops. The new technique is valued highly because it increases crop water-use efficiency and, in accord with increasing public pressure, decreases the allocations of water resources to irrigated agriculture.

Waterlogging and Salinity

Although irrigated areas are extended 5 Mha annually, the net increase is only limited to 2 Mha. This is because of the annually abandoned areas of 2–3 Mha due to salinity and drainage problems. Therefore, there has been only marginal increase in overall irrigated areas in recent years. It is saddening to note that huge irrigation investments made from limited national resources are wasted because of salinity and water-logging. Among the causes are excess water applications, seepage from conveyance networks, lack of administrative organizations, and limited budget allocations for timely maintenance and management programs both at farm and system levels of irrigation schemes. In some cases, soil fertility has decreased to levels that are even lower than those before irrigation was introduced.

Excess use of irrigation water can cause a rise in the water table, and the resulting waterlogging problem can become a major constraint to high crop yields. Soil salinity and waterlogging are interrelated and both occur due to lack of drainage and excess irrigation-water application. Salinity problems associated with irrigated agriculture have been recognized for a long time. These problems, however, are often ignored in the planning stage of irrigation projects. This is not because of the lack of technical understanding, but mainly because of populist political considerations and commitments to extend irrigated acreage to larger areas without drainage (which may hardly provide any political gain). The drainage systems are largely installed underground, and therefore they have the least political appeal. The usefulness of the drainage systems is hard to appreciate in the short term, since they function as preventive measures for soil salination that may occur in the future. Of the 11 Mha of irrigated land in 12 different irrigation projects in India, 2 Mha of good farmlands are waterlogged. One of the major problems confronted by farmers of the Indus Plain in Pakistan is the rise of the water table and the resulting increase in soil salinity. Similarly, after completion of the High Aswan Dam in Egypt, 60% of irrigated farmlands began to have salinity and waterlogging problems.

While water is evaporated from the soil surface and is transpired through plants, the salts contained in irrigation water accumulate in the soil profile. In arid and semiarid regions where annual precipitation is not enough to leach accumulated salts to deeper zones of the soil profile, the soil surface becomes covered with a white salt crust. In addition to the actual crop water requirement, a certain fraction of excess water, called the leaching fraction, must be included in the total applied irrigation water. The excess applied water leaches the salts left in the soil profile and thereby prevents the accumulation of salts. However, if the area lacks properly designed drainage systems, the water table soon rises to the soil surface, continuing evaporation takes place, and the salt accumulation speeds up. The lack of drainage in irrigated areas may lead to soil salination and eventually to the loss of soil fertility, if timely preventive measures are not taken.

Soil salinity that develops because of mismanaged farm irrigation systems is called secondary salinity. Of the world's 1500 Mha of total cultivable land, salt-affected soils make up 340 Mha (23%) and sodium-affected alkaline soils make up 560 Mha (37%). The loss of soil fertility due to soil salinity is a problem that is widespread in nearly all continents. Of irrigated areas, 50% in Uzbekistan, 30% in Brazil and Mexico, 27% in Egypt, 26% in Pakistan, and 13% in China are salt-affected soils with continuing declines of soil fertility.

In many instances, the natural drainage rate may be sufficient to prevent salt accumulation. In the absence of natural drainage, manmade drainage must be installed. In either case, it is essential to have a good drainage outlet to remove saline seeps from agricultural areas. In most natural systems, drainage from upland agricultural areas finds its way into rivers and then to the seas. There may be, however, closed basins without natural outlets. The Dead Sea in the Jordan Valley, the Great Salt Lake in the USA, and the Aral Lake in Central Asia are examples of closed basins storing saline seeps from surrounding irrigated farmlands.

Pollutants Attributable to Irrigation

Sediment load carried in returning waters from large-scale irrigation projects and erosion occurring in watersheds may be a source of major pollutants and result in serious environmental consequences. There may be two different sources of soil erosion: (1) erosion occurring during floods that silts water reservoirs, (2) erosion in poorly planned irrigation project areas where land capability classifications are ignored and surface irrigation methods are practiced with no land grading. Sediments originating from the erosion of fertile surface soil in agricultural areas are carried to downstream areas in returning irrigation waters.

Water reservoirs, manmade small lakes, and large dams may be filled with silt load before they complete their economic life, if forestation is not planned and implemented. Silt and sediment found in water reservoirs may carry pesticide residues (with serious health-hazard effects) and plant nutrients (that may cause algae blooms). Heavy metals are also carried into water reservoirs by way of soil erosion.

Plant nutrients Use of fertilizers is one means of achieving high crop yields under irrigation. However, excess or improper use of fertilizers may pollute air, water, and soil resources. A high percentage of applied fertilizer remains in the soil after the harvest of crops, although this varies depending on the fertilizer practice that is followed, crop species, rainfall, and irrigation management.

With the introduction of new crop cultivars, starting in the 1960s, that showed good response to fertilizer applications, the use of chemical fertilizers showed a significant increase (Figure 3), particularly in Europe and North America. The share of fertilizer consumption in the developing world increased from 20% in the 1970s to 43% in the 1990s. Starting in the early 1980s, fertilizer use no longer showed an increasing trend in America and Europe, with growing public awareness about pollution and degradation of water resources.

Over 60% of applied nitrogen fertilizer may remain in the soil after the harvest of crops. The trend of increasing nitrate and phosphorus concentrations in the watersheds of rivers in South and

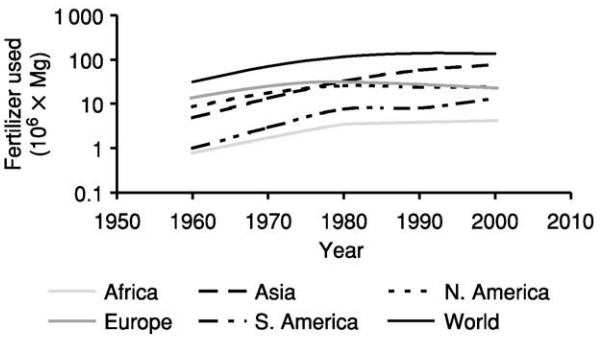


Figure 3 Annual fertilizer consumption by continent (1960–2000). (Adapted from FAOSTAT–Agricultural Data, FAO on-line data base.)

Central Asia since the mid-1970s is attributed to increasing fertilizer usage. Although excess use of potassium may not pose a pollution threat, nitrogen fertilizers can cause pollution in groundwater and surface water resources if care is not taken in their use. Nitrate denitrification occurring in waterlogged areas produces nitrous oxides and other forms of volatilization products that contribute to the atmospheric greenhouse gases responsible for global warming.

Nitrate fertilizers have high solubility, and the end-product nitrate ion is easily leached due to its high mobility with excess irrigation-water application. Hence, nitrate fertilizers mix with groundwater sources. Returning waters from irrigated areas with high sediment loads may also contain high nitrate concentrations. Preventive measures for soil erosion in agricultural areas help to minimize nitrate and phosphorus concentrations in rivers and lakes. The increase in nitrate and phosphorus in lakes is the main cause of eutrophication, algal blooms, and depletion of oxygen that in turn lead to mass fish die-offs. High nitrate concentration in drinking water is also a public health threat. While a high nitrate concentration in drinking water prevents fattening of cattle, it may cause a deadly health disorder called methemoglobinemia in humans (particularly infants). The maximum allowed nitrate concentration in drinking water has been set at $100\,\mathrm{mg\,l^{-1}}$ by the World Health Organization. The European Union restricted it at an even lower limit of $50\,\mathrm{mg\,l^{-1}}$. Agricultural activities are listed among the major causes of river pollution in the USA.

Pesticides Evaporation from large bodies of water stored in dams, networks of channels, and large irrigated fields contributes to increasing humidity, which in turn creates a favorable environment for the spread of plant diseases and pests. Growers, understandably, prefer crops of high cash value under irrigated agriculture. Therefore, it is not uncommon that crop rotation is completely ignored, with the consequences that continuous monoculture of high-value cash crops has become a widespread practice under irrigation.

Pesticides are often heavily applied against weeds and insects under irrigated agriculture. The use of pesticides, in addition to fertilizers, as major inputs of irrigated agriculture dates back to as early as the fifteenth century when arsenic salts were used as herbicides and insecticides. Later, in the sixteenth century, organic pesticides containing nicotine were used as pesticides.

The environmental effects of pesticides were first brought to light in regard to dichlorodiphenyltrichloroethane (DDT), which was first introduced in 1939. The widespread use of DDT continued until the 1970s, when it was banned because of its well-known persistent environmental effects. Despite long years of testing of newly developed pesticides by industry and governmental regulation of their use, pesticides have caused problems such as the development of resistance in target organisms, accumulation in food chains, nonselective killing of all insects, and persistence in the environment.

World pesticide usage exceeded 2.5 million tonnes in 1999. Herbicides accounted for the larger proportion of total usage of 925 000 t, followed by other pesticide usage such as fumigants, nematicides, rodenticides of 756 000 t. Insecticide and fungicide usages were 643 000 t and 252 000 t, respectively. While the share of pesticide use in the developing countries was only 20% in the 1980s, there is an increasing trend in Asia and Latin America, which started in the 1990s. With increased public awareness of health hazards of pesticide residues, they are now more sparingly used in the developed countries and therefore their global use is no longer increasing.

The contribution of pesticide use to overall crop production increase has been estimated to be about 20%. It is therefore unrealistic to assume that their use will cease or be significantly reduced in the future. There should be, however, a rational basis for determining at what point the cost of pesticides becomes prohibitive in economic and environmental terms. Only a small fraction of applied pesticides affects the target organisms and the rest often remains in the soil, contaminates surface water and groundwater resources, or mixes with the air.

Irrigation practices play a major role in the accumulation of pesticide residues encountered in water resources. Residues remaining in the soil are carried into water resources through the leaching of the soil profile with irrigation water. Vaporized pesticide residues mixed in the atmosphere may be carried long distances. The pesticide residues found in arctic mammals were identical to some pesticides used in the tropics and in subtropical areas. Adverse environmental effects of pesticides are further amplified when there is lack of proper training of agricultural labor in the handling and use of pesticides. Death tolls resulting from acute pesticide poisoning are at alarming levels among agricultural labor. It is a widespread problem in Southeast Asia, including Indonesia, Malaysia, and Thailand, where large amounts of pesticides are used in paddy fields. In 1986 in Sri Lanka, 66% of the reported poisoning deaths resulted from the wrong use or mishandling of pesticides.

Heavy metals Fertilizers and pesticides that are used in irrigated agriculture contain a wide array of heavy metals, including arsenic, boron, mercury, copper, cadmium, lead, nickel, and zinc. While boron is often added to the soil with irrigation water, arsenic is included in many insecticides and herbicides. Similarly, many fungicides contain mercury, copper, and cadmium. Phosphorus fertilizers may contain large amounts of heavy metals; they are the major source of cadmium pollution in soils. Heavy metals tend to be adsorbed by clay minerals in the soil profile and therefore are consequently less likely to be leached and mixed with groundwater. However, runoff and tail waters carrying sediments from irrigated areas often contain heavy metals that are discharged into drainage channels and then mixed with other surface water resources. For example, selenium pollution in the San Francisco Delta area shows how sediments carried away from agricultural areas may cause serious environmental damage.

The use of municipal and industrial wastewater in irrigation is another source of heavy-metal pollution in agricultural areas. The discharge of 1 m^3 of untreated municipal or industrial wastewater can contaminate 8–10 m^3 of the clean water source. Hence, the potential for the pollution of water resources is large if one considers that the annual wastewater discharges in Asia, North America, Europe, and Africa are 590, 431, 326, and 55 km^3, respectively.

Public health concerns Socioeconomic benefits and improvement of agricultural infrastructure are major gains that may result from the implementation of large-scale irrigation projects. However, immediately following irrigation development projects, there may be major demographic changes as a high number of people from rural areas move to cities or agrobusiness centers to take advantage of emerging employment opportunities. There may also be undesirable changes such as an increase in pests and parasites carrying diseases, related to the large bodies of water stored in dams and water-conveying open-channel systems and related climate changes in the project area.

Among the diseases spread by parasites and insects feeding in water sources are malaria, schistosomiasis (bilharzias), yellow fever, brain fever (Japanese encephalitis), and river blindness. Brain fever is largely found in the rice fields of South, Southeast, and West Asia; lymphatic filariasis, in all continents; river blindness, largely in Africa and Central America; and malaria and schistosomiasis, in all continents (but largely in Africa). At the beginning of the twenty-first century, two million people die each year because of malaria. Following the construction of the Diama dam in Senegal, there was a sudden burst of schistosomiasis incidence. Similarly, the number of schistosomiasis patients increased threefold after the completion of the Aswan dam. Malaria incidence has increased because of seepage and waterlogging in the irrigated plains of Punjab and Haryana, in India. Similarly, in Latin America, as much as a threefold increase in malaria incidence was noted during a short span of time following implementation of open-channel irrigation (Figure 4).

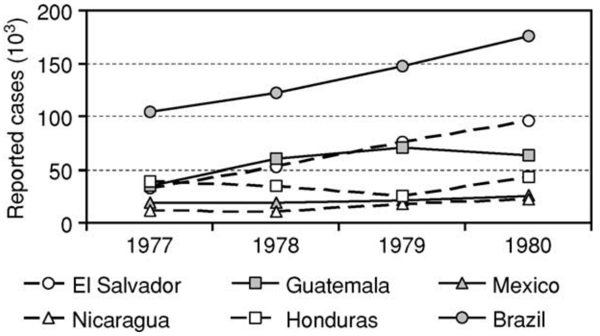


Figure 4 Incidence of malaria due to introduction of canal irrigation in some Latin American countries. (Adapted from WHO (1983) *Report of the Experts on Environmental Management for Vector Control*. VBC/83.4. Geneva, Switzerland: WHO.)

Municipal and industrial wastewaters are often recycled and rerouted into irrigation schemes in water-scarce regions. Although treatment of such water sources is common practice in the developed world, examples of water treatment in developing countries are rare. Wastewaters from poultry farms, cattle ranges, and household sewage may pose a serious health hazard if used for the irrigation of freshly consumed vegetables. Among the diseases spread because of the use of untreated wastewater for irrigation are cholera, typhoid, ascariasis, amebiasis, giardiasis, and enteroinvasive *Escherichia coli*. Five million people die annually from diseases related to the practice of irrigation with untreated wastewater.

Sustainability

The history of agriculture has shown that sustainable irrigated agriculture requires adequate salt balance and drainage. The technology to accomplish this exists, but its use is limited by numerous factors. These factors include the lack of economic incentives for growers to improve their existing irrigation structures; need for investment in pressured irrigation systems and drainage works; lack of training on and experience with effective management practices; public opposition to the possible effects of irrigation return flows on the environment; and administrative constraints such as low budget allocations for timely maintenance and management of irrigation schemes.

Recent developments in irrigation technology are helping to improve irrigation efficiency and to reduce the drainage water volume that requires disposal. For example, sprinkler and drip irrigation systems facilitate the fine-tuned control of fertilizer applications in order to minimize fertilizer residue remaining in the soil after the harvest of crops. The new irrigation technologies that facilitate the application of fertilizers along with irrigation water should be adopted and used widely. Irrigation water requirements are least, for example, with drip systems that supply water and fertilizers directly into the root zone where they are needed. The crop utilizes nearly all of the fertilizer applied, leaving the least residue in the soil after harvest.

Over-use of pesticides is now being replaced with integrated pest management practices that take advantage of the biological control of insects and pests in addition to the use of chemical sprays. However, efforts are still needed to create public awareness of the importance of conservation of soil and water for meeting the needs of present and future generations.

Training farm advisers and extension workers on effective management practices of land preparation and soil tillage, and use of pesticides and fertilizers will help to minimize pollution attributable to agricultural activities. Adopting advanced irrigation technologies supported by governmental subsidies or low-interest loans (if needed), and drafting and enforcing legal codes for preventing pollution of water and soil resources will contribute to the environment-friendly and sustainable practice of irrigated agriculture. Presently available technical knowledge on irrigation is adequate for developing sustainable agricultural practices to keep pace with the food and fiber demands of the increasing world population. To this effect, despite the drawbacks, irrigated agriculture has an important role to play in world food production.

See also: **Civilization, Role of Soils**; **Drainage, Surface and Subsurface**; **Salination Processes**; **Salinity:** Management; **Salt-Affected Soils, Reclamation**

Further Reading

Alexandratos N (ed.) (1995) *World Agriculture: Towards 2010. An FAO Study.* Chichester, UK: FAO/John Wiley.

Barrios A (2000) *Agriculture and Water Quality.* CAE Working Paper Series, WP00-2. DeKalb, IL: American Farmland Trust's Center for Agriculture in the Environment.

Daugherty TC and Hall AW (1995) *Environmental Impact Assessment of Irrigation and Drainage Projects.* FAO Irrigation and Drainage Paper 53. Rome, Italy: FAO.

Donaldson D, Kiely Timothy, and Grube A (2002) *Pesticides Industry Sales and Usage. 1998 and 1999 Market Estimates.* Biological and Economic Analysis Division Office of Pesticide Programs, Office of Prevention, Pesticides, and Toxic Substances, US Environmental Protection Agency. Washington, DC: US Government Printing Office.

FAO (2002) *Deficit Irrigation Practices.* Water Reports 22. Rome, Italy: FAO.

FAO (2002) *World Agriculture Towards 2015/2030. Summary Report.* Rome, Italy: FAO.

Hornsby AG (1990) Pollution and public health problems related to irrigation. In: Stewart BA and Nielsen DR (eds) *Irrigation of Agricultural Crops*, pp. 1173–1188. Agronomy Monograph No. 30. Madison, WI: American Society of Agronomy.

Kirda C, Moutonnet P, Hera C, and Nielsen DR (eds) (1999) *Crop Yield Response to Deficit Irrigation.* Dordrecht, The Netherlands: Kluwer Academic Publishers.

Marr PD (1969) The social context of irrigation. In: Hagan RM, Haise HR, and Edminster TW (eds) *Irrigation of Agricultural Lands*, pp. 12–22. Agronomy Monograph No. 11. Madison, WI: American Society of Agronomy.

Shah T, Molden D, Sakthivadivel R, and Seckler D (2000) *The Global Groundwater Situation: Overview of Opportunities and Challenges.* Colombo, Sri Lanka: International Water Management Institute.

Shiklomanov LA (1999) *World Water Resources at the Beginning of the 21st Century.* IHP Report. Paris, France: UNESCO.

Tanji KK (ed.) (1990) *Agricultural Salinity Assessment and Management.* Manuals and Reports on Engineering Practice No. 71. New York: American Society of Civil Engineers.

Umali DL (1993) *Irrigation Induced Salinity: A Growing Problem for Development and the Environment.* World Bank Technical Paper 215. Washington, DC: The World Bank.

Wagner RH (1971) *Environment and Man.* New York: Norton.

Methods

D L Bjorneberg and R E Sojka, USDA Agricultural Research Service, Kimberly, ID, USA

Published by Elsevier Ltd.

Introduction

Irrigation is the process of applying water to soil, primarily to meet the water needs of growing plants. Water from rivers, reservoirs, lakes, or aquifers is pumped or flows by gravity through pipes, canals, ditches, or even natural streams. Applying water to fields enhances the magnitude, quality, and reliability of crop production – approximately 30% of the world's food is grown on irrigated land, which accounts for only about 15% of the world's land used for crop production.

Various irrigation methods have been developed over time to meet the irrigation needs of certain crops in specific areas. The three main methods of irrigation are surface, sprinkler, and drip or microirrigation. In surface irrigation, water flows over the soil by gravity. Sprinkler irrigation applies water to soil by sprinkling or spraying water droplets from fixed or moving systems. Microirrigation applies frequent, small applications by dripping, bubbling, or spraying. A fourth, and minor, irrigation method is subirrigation, where the water table is raised to or held near the plant root zone using ditches or subsurface drains to supply the water.

More than 90% of the approximately 270 Mha of irrigated crop land in the world is surface-irrigated; less than 1% is irrigated by microirrigation. In the USA, sprinkler and microirrigation are used on a greater percentage of irrigated cropland, primarily to reduce labor and improve control of water application. Approximately 45% of the 20 Mha of cropland irrigated in the USA is sprinkler-irrigated, and microirrigation is used on approximately 4% of the irrigated cropland.

Surface Irrigation

Surface irrigation entails water flowing by gravity over soil. Water is usually supplied by gravity through canals, pipes, or ditches from the water source to the field. In some locations, however, water may need to be pumped from the source to a field at a higher elevation. Types of surface irrigation systems include furrow, basin, and border irrigation. Surface irrigation systems are typically used for field crops, pastures, and orchards. Efficiency of surface irrigation systems varies tremendously because of variations in soil type, field uniformity, and management. Surface irrigation is often considered less efficient than sprinkler irrigation or microirrigation, because soil, not a pipe, is the water-conveyance system for surface irrigation (Table 1). However, a well-managed surface irrigation system on a uniform soil with a runoff-reuse system can approach 90% application efficiency.

Table 1 Application efficiencies for irrigation systems

System type	*Application efficiency (%)*
Surface irrigation[a]	
Furrow	50–70
Level basin	60–80
Border	60–75
Sprinkler irrigation	
Solid set	60–85
Set move	60–75
Moving[b]	75–95
Traveling gun	55–65
Microirrigation[c]	80–95
Subirrigation	50–80

[a] Surface irrigation efficiencies can be greater if runoff is reused.
[b] Includes center-pivot, linear move, and low-energy precision application (LEPA) systems.
[c] Efficiency can decrease to 50% with poor management.

Furrow Irrigation

In furrow irrigation, water flows in evenly spaced furrows or corrugates that are typically 0.1–0.3 m wide (Figure 1) on fields with slopes of 0.1–3%. Water commonly flows in furrows for 12–24 h during an irrigation, but shorter or longer durations may be used depending on furrow length and soil, water, and management considerations. Inflow rates for individual furrows vary from approximately 10 to 100 l min^{-1}, again depending on soil, slope, field length, and management. Ideally, water should advance across the field in approximately 25% of the total irrigation time to irrigate the field uniformly. Since soil erosion increases as field slope and inflow rate increase, flow rate must be carefully managed on fields with steeper slopes (greater than 1%). Low inflow rates and long irrigation durations may be needed to apply the desired amount of water during

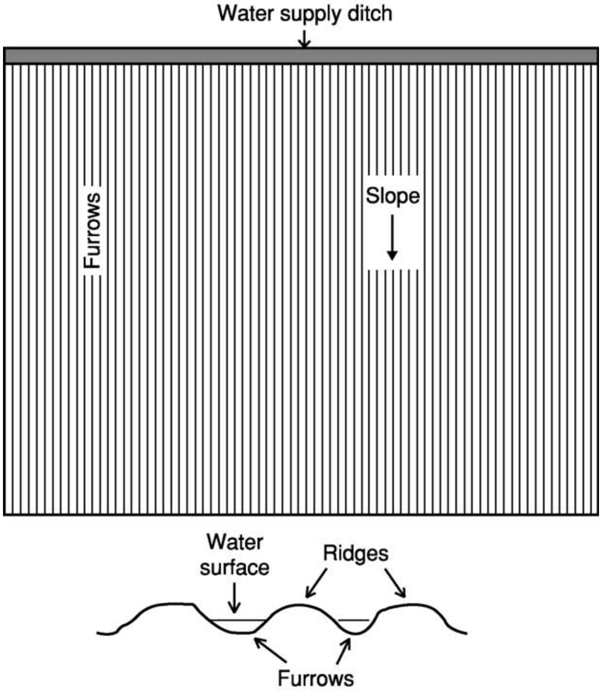


Figure 1 Schematic of a furrow-irrigated field, where water flows in evenly spaced furrows or corrugates.

Figure 2 Furrow-irrigated sugar beet field, with water supplied by siphon tubes from a concrete ditch.

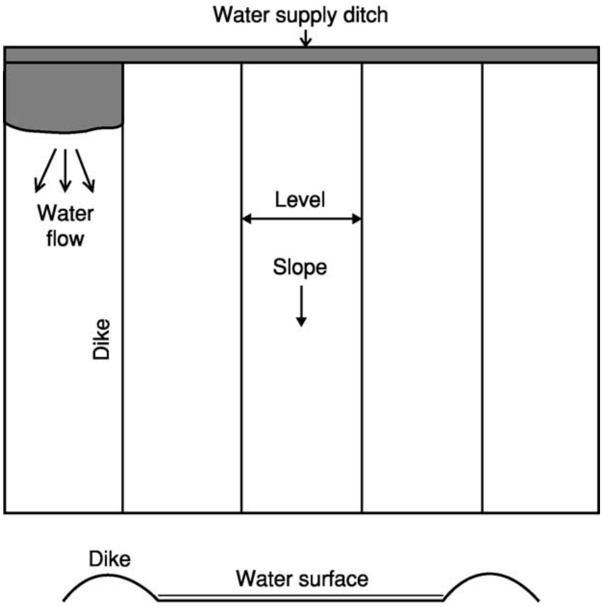


Figure 3 Schematic of a border-irrigated field, where a uniform sheet of water flows between dikes.

an irrigation on soils with low infiltration rate. Conversely, higher inflow rates are often needed on fields with low slopes and/or high infiltration rates in order for the water to flow across the field and uniformly irrigate the upper and lower ends of the field.

Inflow to irrigation furrows may be supplied from gated pipes or ditches (earthen or concrete). Siphon tubes are frequently used to convey and regulate water flow from ditches to individual furrows (Figure 2). By creating a siphon, water flows through the tube, over the ditch bank, and into the furrow, as long as the tube outlet is lower than the water elevation in the ditch. Furrow inflow rate is controlled by tube diameter and the elevation difference between the ditch-water level and tube outlet. Gated pipes distribute water to furrows through evenly spaced outlets on the pipes. With earthen ditches, water can flow through a breach or other opening in the ditch bank to individual furrows or a smaller feed ditch that distributes water to several furrows. It is much more difficult to regulate flow through a breach in an earthen ditch than through siphon tubes or pipe gates.

Furrow irrigation requires lower capital investment, less technical knowledge, and greater labor than most other irrigation systems. Fields can be irrigated without leveling or grading, because water flows in furrows. Furrow irrigation is not well suited to automation, because water flow rate must be adjusted for each furrow for each irrigation.

Basin and Border Irrigation

Basin and border irrigation systems are similar in that both involve a uniform sheet of water flowing over the soil. The general difference is that basin irrigation involves applying water to a nearly level field and may include ponding for extended time periods. With border irrigation, water flows between dikes that divide a sloping field into rectangular strips with free drainage at the end (Figure 3). The purpose of the dikes is to contain water as it flows across the field, unlike basin irrigation where the dikes pond the water.

Basins can be as small as a few square meters for a single tree or as large as several hectares with greater than $100\,\mathrm{l\,s^{-1}}$ inflow rates. Basin size is a balance between soil infiltration rate, slope, and water supply. Water depth in basins varies from approximately 5 to 20 cm, with typical depths of 10–15 cm. Efficient basin irrigation requires a level soil surface with uniform soil texture and adequate water supply so the basin is quickly and uniformly covered with water. If the basin is not level, the higher elevation areas will receive less water than the low areas. If the basin inflow rate is inadequate, water will slowly advance, causing large differences in infiltration opportunity time within the basin.

A special type of basin irrigation is a drain-back level basin. Drain-back level basins have a series of parallel basins that receive inflow from a shallow, 5–10-m-wide ditch. After the first basin is filled, a gate opens to start filling the adjacent basin, which is at a lower elevation. Water near the inflow end of the first basin drains back to the inflow ditch and flows to the next basin. This procedure is repeated until every basin has been irrigated. The drain-back phase improves uniformity by reducing the amount of water that infiltrates near the inflow end and initially increases the inflow rate to the next basin, which increases the advance rate.

Border irrigation systems are better suited for sloping fields than basin systems, because water flows between dikes rather than ponds within basins. The irrigated areas between dikes can be 3–30 m wide and up to 400 m long. The field slope between dikes (perpendicular to water flow direction) should be nearly level so water flows uniformly down the field. The slope along the dikes can be similar to furrow irrigation, but border systems often have slopes of less than 0.5%.

Water can be supplied to borders and basins from open ditches with gates, breaches, or siphon tubes or from above- or belowground pipes. Typical inflow rates vary from 10 to $100\,\mathrm{l\,s^{-1}}$, but vary widely depending on the size of basin or border, soil texture, and slope. Border and basin irrigation require less labor than furrow irrigation because water is supplied to a larger area with a single outlet.

Sprinkler Irrigation

Sprinkler irrigation applies water to soil by spraying or sprinkling water through the air on to the soil surface. Water is pressurized and delivered to the irrigation system by a mainline pipe, which is often buried so it does not interfere with farming operations. Three main categories of sprinkler irrigation systems are solid-set, set-move, and moving. Sprinkler irrigation is used for a wide variety of plants, including field crops, orchard trees, turf, and pasture. Sprinkler systems are also installed for applying wastewater and protecting plants from frost.

Solid-set systems may be installed for a single season for certain field crops or permanently for turf, orchards, or permanent crops. Set-move systems are manually or mechanically moved to another part of the field after the irrigation set is complete in the present location. Moving systems such as center pivots or traveling guns apply water as the system slowly travels through the field.

Sprinkler irrigation is often more efficient than surface irrigation, because water application is more controlled (Table 1). In hot and/or windy areas, however, sprinkler irrigation can have significant losses to evaporation and wind drift. Maintenance is also important for efficient sprinkler irrigation. Worn nozzles and leaking pipe connections reduce application uniformity and system efficiency.

Solid-Set Sprinkler Systems

Most solid-set sprinkler irrigation systems are designed to apply frequent, small amounts of water to meet plant water needs (e.g., daily). Water application rates can vary from approximately $4\text{–}6\,\mathrm{mm\,h^{-1}}$ for field crops to $5\text{–}30\,\mathrm{mm\,h^{-1}}$ for turf applications. Overhead costs are greater for solid-set systems than other sprinkler systems, because the entire irrigated area must be equipped with sprinklers and pipe. However, permanently installed systems can be automated to reduce labor and allow irrigation at any hour of the day, which reduces the opportunity for plants to be stressed. When properly designed, solid-set systems have high application uniformity. While solid-set systems are most commonly used with turf, and landscape and permanent crops, these systems are also used for some high-value annual crops with low tolerance for water stress.

Solid-set system designs are as varied as the applications: small sprinklers may irrigate $20\,\mathrm{m^2}$, or large, gun-type sprinklers may be spaced 50 m apart. Plastic pipe is frequently used for buried applications, but it is also used in some aboveground applications. Aluminum pipe (50–100 mm in diameter) is often used for field crops when the system is installed after planting and removed before harvest. Most systems are divided into zones so a portion of the area is irrigated at one time. Solid-set systems used for frost control, however, must be designed to water the entire area simultaneously.

Set-Move Sprinkler Systems

Set-move sprinkler irrigation systems are designed to apply water slowly during the irrigation set (e.g., $4\text{–}6\,\mathrm{mm\,h^{-1}}$), which often lasts 8–24 h. After completing the irrigation set, the sprinkler system is moved to an adjacent area for the next set (Figure 4). Adequate water has to be applied during an irrigation set to meet the crop's water needs until the system is moved back to the area, often in 7–10 days.

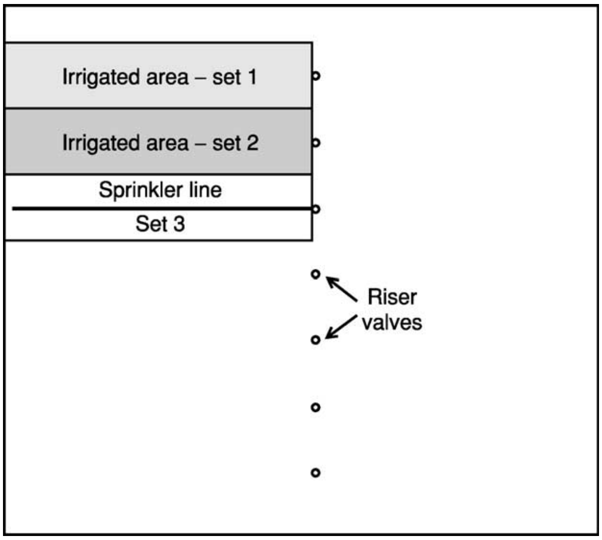


Figure 4 Schematic of a set-move-irrigated field, where the irrigation system is moved to the adjacent area after completing an irrigation set.

The common types of set-move irrigation systems are hand-move and side-roll systems. Hand-move systems can be a single sprinkler or a line of sprinklers. A line of hand-move sprinklers, sometimes called 'handlines,' is typically composed of 9- or 12-m-long pieces of 75- or 100-mm-diameter aluminum pipe with a sprinkler mounted on one end or in the center. Individual pipes are connected to form an irrigation line, usually not more than 400 m long. After an irrigation set is completed, the line is disconnected, and each piece is moved by hand 10–20 m to the next set. A slight variation to the handline is the dragline or end-pull system. These systems, which are less common, have special connections between sprinkler pipes that allow the irrigation line to be pulled by a tractor to the next set.

Side-roll systems, also called wheel lines, are similar in principle to handlines except a large-diameter wheel (1.5–3 m in diameter) is mounted in the center or on the end of each piece of aluminum pipe (100–125 mm in diameter) to elevate the sprinkler. The sprinkler pipe is the axle for the side-roll. When an irrigation set is completed and the pipe has drained, the wheel line, powered by an engine, is rolled to the next position. Self-leveling sprinklers are used so the side-roll does not have to be exactly positioned for the sprinklers to operate correctly.

Moving Sprinkler Systems

Moving irrigation systems include center-pivot, linear-move, and traveling gun systems. Center-pivot and linear-move systems are similar in design and appearance. These systems consist of one or more spans of sprinkler pipe elevated by 'A-frame' towers. Span length varies from 30 to 65 m. Towers, powered by hydraulic or electric motors, elevate the sprinkler pipe 2–4 m above the ground.

The center pivot has a stationary pivot point so the towers move in a circle (Figure 5). Water and power are supplied to the system through the pivot point. A typical center pivot in the USA has eight spans, a total length of approximately 400 m, and irrigates 50–60 ha. Center pivots are extremely popular, because water is uniformly applied to a large area with little labor. Furthermore, once a circular field has been irrigated, the center pivot is in position to start the next irrigation. System cost per irrigated area is reduced by increasing the total length of a center pivot, because irrigated area per unit length increases with distance from the pivot point. Consequently, water application rate increases with distance from the pivot point, because each span must irrigate a larger area per revolution (a 50-m span at the pivot point irrigates 0.8 ha, while a 50-m span that is 350 m from the pivot point irrigates 12 ha). Application rates often exceed the infiltration rate of the soil under the outer spans of center pivots. Thus, the opportunity for runoff increases as center-pivot length increases. Center pivots can also irrigate fields with rolling terrain that are difficult or impossible to irrigate by surface-irrigation methods; however, these conditions can create additional management challenges.

Linear-move systems have a control unit on one end, or in the center on longer systems, that moves the towers in a straight line to irrigate rectangular-shaped fields. Power is supplied by an electrical drag cord or by an engine-powered generator mounted on the control unit. Water is typically supplied to the drive unit by a drag hose connected to a buried or aboveground pipe. Drive units can be equipped with a pump so water can be supplied from an open ditch flowing parallel to the travel direction, but this is not commonly used, because it requires a nearly level ditch. System cost per irrigated area is reduced by increasing the distance the system travels. Since hose length is limited to approximately 150 m, the system can move 300 m before the hose is connected to the next riser. Similar to set-move systems, adequate water must be applied to meet crop needs until the linear-move can irrigate the area again.

Early center pivots had impact sprinklers mounted on top of the irrigation pipe that required 500–600 kPa. Most new systems have low-pressure nozzles (70–200 kPa) mounted on tubes that extend below the

Figure 5 Center-pivot irrigation system.

irrigation pipe, so nozzle height varies from 1 to 3 m above the soil. Manufacturers make numerous types of low-pressure nozzles with fixed or spinning spray plates that provide a wide range of application rates and water-droplet sizes to meet field conditions and operator preferences. A common feature to all nozzle types is a pressure regulator, which maintains constant nozzle pressure as the system travels across a field with varied elevation.

A traveling gun has a large-capacity sprinkler on a cart that is pulled across the field by a cable or by the water-supply hose. These systems irrigate an area 50–100 m wide and up to 400 m long. A traveling gun can be considered a moving, set-move system because water is applied as the cart moves across the field and then the system is moved to another area in the field for the next irrigation set. For cable-tow systems, a winch on the cart winds the cable, pulling the cart and a soft hose across the field. A hose-reel system pulls the cart as a hard plastic (polyethylene) hose is wound around a reel on a trailer anchored at the end of the run. The reel or winch is powered by an engine or a water turbine. Smaller versions of traveling guns are available for irrigating athletics fields, small pastures, or arenas.

Microirrigation

Microirrigation applies water at low rates and pressures to discrete areas so irrigation water reaches the root zone with minimal losses. Water drips from emitters in plastic pipe or tape, or bubbles or sprays from small emitters that only wet a portion of the soil surface. Microirrigation systems are popular for permanently installed systems that irrigate trees, vineyards, orchards, and shrubs. These systems are typically automated so that water is applied frequently (e.g., daily) to maintain optimum soil-water content near the plants. Filtration is important for microirrigation, because sediment and algae can plug the small openings on drip emitters, bubblers, and microsprays. Chemical treatment may also be necessary to reduce salt or mineral deposits that can plug emitters.

Drip irrigation emitters are preinstalled within polyethylene pipe at regular intervals or emitters are attached to the outside of the pipe at desired locations. Emitter flow rates typically vary from 2 to 7.5 $\mathrm{l\,h^{-1}}$. Pressure-compensating emitters maintain a constant flow rate as pressure varies from approximately 70 to 200 kPa (Figure 6). This type of system is common in vineyards.

Figure 6 Pressure-compensating drip emitter watering ornamental plants.

Figure 7 Drip tape installed on the soil surface to irrigate edible beans.

Drip tape is thin-walled (0.1–0.375 mm) plastic tubing (10–20 mm in diameter), with outlets at 100–600-mm intervals (Figure 7). Flow rates can vary from 100 to 400 $\mathrm{l\,h^{-1}}$ per 100 m of length. Typical operating pressure for drip tape is 35–100 kPa. Drip tape is commonly installed below the soil surface, where there is less opportunity for damage. Buried drip tape can be used for several seasons or retrieved after a single season, depending on crop types and farming practices.

Bubblers and microsprays are often used for irrigating trees, shrubs, or ornamental plants. Bubblers discharge water with low energy to flood a small area. Flow rates up to 100 $\mathrm{l\,h^{-1}}$ can apply water to 4-m-diameter areas, depending on nozzle size, type, and pressure. Microsprays apply a fine spray or mist with similar flow rates and wetted areas as bubblers.

Subirrigation

Subirrigation applies water below the soil surface to raise the water table into or near the plant root

Table 2 Typical advantages and disadvantages of irrigation systems

System type	*Advantages*	*Disadvantages*
Surface irrigation		
Furrow	Low capital and maintenance costs; water flows in small channels	High labor; less water control; soil erosion; possible runoff and percolation losses
Level basin	Efficient with good design; less labor than furrow	Ponded water; sloping fields must be leveled
Border	Less labor and less runoff than furrow; easier to manage infiltration depth	Water flows over entire soil surface
Sprinkler irrigation		
Solid set	Good water control; possible to automate and frequently irrigate; fits odd-shaped fields	High capital cost; system may interfere with field operations
Set-move	Lower capital cost than other sprinkler systems	More labor than other sprinkler systems; poor uniformity in windy conditions; greater application depth
Moving[a]	High uniformity; low labor	High capital and maintenance costs; not suitable for odd-shaped fields; potential wind and evaporation losses
Traveling gun	Lower capital cost than other sprinkler systems	Higher operating cost; wind and evaporation losses
Microirrigation	Excellent water control; frequent applications possible	Higher capital cost; requires clean water or treatment and filtration

[a] Includes center-pivot, linear-move, and low-energy precision application systems.

zone. Subirrigation is not often used in arid or semiarid irrigated areas; it is typically used in conjunction with subsurface drainage. Subsurface drainage lowers the water table and removes excess water through open ditches or perforated pipe. Water-table depth can be controlled by installing a weir on the drainage system. During wet periods, the water table is lowered so the root zone remains unsaturated. During dry periods, water is pumped into the drainage system to raise the water table and provide additional water for plant growth. In some situations, drained water can be stored for use when irrigating.

Salinity Hazards

Salinity problems are more closely associated with soil and water chemistry than the type of irrigation system; however, the irrigation system can accentuate salinity problems and affect how salinity is managed. Basin and border irrigation can uniformly leach salts from the soil when infiltration is uniform. Sprinkler irrigation can also uniformly leach salts, but sprinkling can injure sensitive plants when highly saline water is applied to foliage. Furrow irrigation leaches salt from soil below the furrow, while increasing salt concentrations in the bed. Tilling before planting crops tends to minimize salinity problems by mixing salts in the soil. Salt concentration tends to increase radially from a drip emitter or laterally from a line source such as drip tape. Well-managed drip irrigation can minimize salt-induced stress by maintaining a high soil-water content with frequent irrigations.

Choosing an Irrigation System

Choosing an irrigation system is a difficult task. Irrigation systems are as varied as the people who use them. The right selection for a user depends on soil, water, and climatic conditions, as well as crop types, user knowledge and preference, capital and operating costs, and infrastructure availability. No system is best for all situations. Some typical advantages and disadvantages of irrigation systems are shown in Table 2. Sprinkler or microirrigation are often better choices than surface irrigation on sandy soil, where excessive percolation is a problem. Surface irrigation may be better in arid, windy areas, where wind and evaporation losses can be significant. Surface irrigation offers less control of application depth, so small, frequent irrigations are not practical for water-sensitive crops, which are better suited to microirrigation, solid-set, or center-pivot systems.

List of Technical Nomenclature

ha	hectares
kPa	kilopascals
$\mathrm{l\,h^{-1}}$	liters per hour
$\mathrm{l\,min^{-1}}$	liters per minute
$\mathrm{l\,s^{-1}}$	liters per second
Mha	million hectares
m	meters
$\mathrm{m^2}$	square meters
mm	millimeters

See also: **Crop Water Requirements**; **Evapotranspiration**; **Fertigation**; **Infiltration**; **Irrigation:** Environmental Effects; **Salinity:** Management; **Water Harvesting**; **Water-Use Efficiency**

Further Reading

Brouwer C, Prins K, Kay M, and Heibloem M (1988) *Irrigation Water Management: Irrigation Methods*. Training manual No. 5. Rome, Italy: Food and Agriculture Organization of the UN.

Cuenca RH (1989) *Irrigation System Design: An Engineering Approach*. Englewood Cliffs, NJ: Prentice-Hall.

FAO *Aquastat – Information System on Water and Agriculture*. Rome, Italy: Food and Agriculture Organization of the UN.

Hoffman GJ, Howell TA, and Solomon KH (1990) *Management of Farm Irrigation Systems*. St. Joseph, MI: American Society of Agricultural Engineers.

Jensen ME (1983) *Design and Operation of Farm Irrigation Systems*. ASAE Monograph No. 3. St. Joseph, MI: American Society of Agricultural Engineers.

USDA (1997) *1997 Census of Agriculture*, vol. 1. *National, State, and County Tables*. National Agricultural Statistics Service. Washington, DC: US Government Printing Office.

ISOTOPES IN SOIL AND PLANT INVESTIGATIONS

K Reichardt and O O S Bacchi, Center for Nuclear Energy in Agriculture, Piracicaba, Brazil

Introduction

Isotopes, both radioactive and stable, have been used successfully as tracers in environmental studies involving the soil–plant–atmosphere system. Each chemical element of the periodic table has isotopes, which are atoms of the same species, behaving chemically and biologically in the same manner, differing only in some of their physical properties. They have in their nuclei the same number of protons, which defines the element, but different number of neutrons, which makes them different in terms of mass. Most of them are natural, i.e., present in nature since the formation of the Earth, and some are artificial, laboratory-made through nuclear reactions. When stable, they differ only by their atomic weight, and, when radioactive (unstable), they emit radiation. The important thing is that they differ in a measurable way and can be detected. Stable isotopes are detected by mass spectrometry, with instruments capable of distinguishing atomic weights, and radioisotopes, by radiation detection, made by a wide range of detectors, depending on the type of the radiation. One important feature of these methodologies is that they are able to measure extremely low amounts of isotopes.

Radioisotopes as Tracers

Radioisotopes are unstable isotopes that approach stability by emitting radiation. This radiation can easily be detected even in extremely low amounts. Radioisotopes can be traced at concentrations as low as 10^{-11} (1/100 billion). Each radioisotope has its own characteristics and emits one or more radiation types (mainly α (alpha), β^- (beta negative), β^+ (beta positive), γ (gamma), and n (neutron)), with one or more energies, and at a given rate that depends on their half-lives. Since this rate decreases exponentially with time, the half-life T is defined as the time in which any emission rate decreases by half. The type and energy of radiation affect detection of radioisotopes and a short half-life can limit their use in long-term experiments.

Radioisotopes can be found naturally, for example, ^{3}H, ^{14}C, ^{40}K, ^{226}Ra, and ^{235}U, some of them being continuously produced by natural processes occurring in the upper atmosphere which are induced by solar or cosmic radiation. In this case the production rate of the radioisotope is in equilibrium with its decay rate. Some radioisotopes have half-lives longer than the age of our planet, therefore they are present naturally in the environment. The great majority of radioisotopes used as tracers are, however, artificially produced in nuclear reactors. Table 1 presents the principal tracer isotopes used in environmental studies, mainly in plant biology and soil studies.

To use isotopes as tracers, they are added to the respective stable isotope in amounts that allow their detection to the end of the experiment. Whenever possible they should be added in the same chemical form as the stable compound, e.g., if a phosphorus study is performed using superphosphate, the radioactive ^{32}P has to be present also in the form of superphosphate. Sometimes this is costly, as is the case in herbicide studies in which the radioactive ^{14}C has to be incorporated into the same molecular species.

Table 1 Some tracer radioisotopes for soil, plant, and environmental studies

Most abundant isotope	*Tracer radioisotope*	*Radiation energy (MeV)*		*Half-life*
		Beta	*Gamma*	
Calcium-40	^{45}Ca	0.254		153 days
Carbon-12	^{14}C	0.156		5720 years
Cesium-133	^{137}Cs	0.52; 1.18	0.662	30 years
Cobalt-59	^{60}Co	0.31	1.17; 1.33	5.27 years
Hydrogen-1	^{3}H	0.0181		12.26 years
Iodine-127	^{131}I	0.61; 0.25; 0.85	0.36; 0.08; 0.72	8.05 years
Magnesium-24	^{28}Mg	0.45	0.032; 1.35; 0.95	21.3 h
Manganese-55	^{52}Mn	0.6	0.94; 1.46	5.7 days
Phosphorus-31	^{32}P	1.71		14.3 days
Potassium-39	^{40}K	1.32	1.46	1.3×10^{9} years
Rubidium-85	^{86}Rb	0.7; 1.77	1.08	18.7 days
Strontium-88	^{89}Sr	1.47		50.4 days
Sulphur-32	^{35}S	0.168		86.7 days
Zinc-64	^{65}Zn	0.33	1.11	245 days

The strength of radioactive materials is measured in terms of activity ($1\,\text{Beq} = 1$ disintegration s^{-1}; $1\,\text{Ci} = 3.7 \times 10^{10}$ disintegrations s^{-1}). Each detected radiation corresponds to one count, and therefore activities are also expressed as counts per second (cps) or counts per minute (cpm). Since the measured activity depends on the size of the sample, the specific activity (S) is defined as the activity per unit of mass, typically expressed as cps per milligram or cpm per microgram.

From the above it can be seen that radioisotopes can be used as tracers in a wide range of experiments. Sometimes one looks only for the presence of the isotope in a given part of the system, and in other situations one needs to measure amounts absorbed or transferred, or rates of specific processes. The basis of most applications lies in the principle of isotope dilution, which states that, for a given constant amount of radioactivity, the specific activity is inversely proportional to the total amount of test element or substance present in the system. The technique has several advantages, being especially useful in obtaining analytical data when chemical separations are difficult, time-consuming, or impossible. It allows the estimation of uptake of a nutrient by a plant from fertilizer, or the recognition and quantitative determination of intermediates in metabolism studies. In a fertilizer study, if a plant receives ^{32}P-labeled superphosphate, a distinction can be made between the soil P and the fertilizer P absorbed by the plant. The percentage phosphorus derived from the fertilizer (%Pdff) can be calculated as follows:

$$\%\,\text{Pdff} = \frac{S \text{ of } ^{32}\text{P in plant material}}{S \text{ of } ^{32}\text{P in fertilizer}} 100 \qquad [1]$$

For example, if the plant material of a crop grown on a soil fertilized with a ^{32}P-labeled fertilizer presents a specific activity S of $35\,758\,\text{cpm}\,\mu\text{mol}^{-1}$ P, and the applied fertilizer has a specific activity of $102\,441\,\text{cpm}\,\mu\text{mol}^{-1}$ P, one can say that approximately 35% of the phosphorus present in the plant comes from the fertilizer. Such experiments are used to study fertilizer-use efficiency, optimization of fertilizer placement and rate, evaluation of varieties, and to answer several other important questions raised during soil–plant management.

Radioactive tracers have been extensively used in studies concerning environmental pollution, dynamics of nutrients in the soil–plant–atmosphere system, the fate of herbicides and other agrochemicals, etc. One important application is the use of fallout ^{137}Cs as a label for soil erosion. The nuclear tests made in the middle of the twentieth century have spread over the whole surface of the Earth a fairly constant amount of ^{137}Cs, an element that is strongly bound to the soil, remaining in the surface layer. When soil is washed downslope, the cesium is carried on soil particles, and in this way it is possible to identify erosion areas in which the ^{137}Cs content is lower than expected, and deposition areas with higher ^{137}Cs content.

Stable Isotopes as Tracers

Radioactive isotopes of suitable half-life do not exist for some elements, in which case stable isotopes may be used. The quality of stable isotopes is defined in terms of their relative abundance. For example, the ^{15}N abundance of a sample containing nitrogen is the ratio of ^{15}N atoms to the total of N atoms in the sample, which in the case of nitrogen, which has only two isotopes, is equal to ($^{15}\text{N} + {}^{14}\text{N}$), i.e.:

$$\%\,^{15}\text{N abundance} = \frac{\text{atoms } ^{15}\text{N}}{\text{atoms } ^{15}\text{N} + \text{atoms } ^{14}\text{N}} 100 \qquad [2]$$

The natural environment, due to the intensive nitrogen-cycling in biological materials over billions of years, has a fairly constant ^{15}N abundance of 0.365%. Therefore, amounts of ^{15}N in enriched samples can conveniently be expressed as percentage of ^{15}N atoms in excess of the natural abundance of 0.365.

Dynamic processes in which the flow of compounds is affected by atomic mass affect the relative abundance of isotopes. This fact is used in the laboratory to isotopically enrich compounds. Using batteries of long resin-exchange columns, ammonium sulfate can be enriched with ^{15}N. In this way, this compound, which is a nitrogen fertilizer, can be enriched with ^{15}N to an abundance of, e.g., 10.365%, having therefore an atom excess of 10.000%. This fertilizer can be used as a nitrogen tracer in agronomic studies. If a soil is fertilized with this compound, a plant growing on this soil will have two sources from where to draw nitrogen, the native soil N, with an abundance of 0.365%, and the fertilizer N, with an abundance of 10.365%. The tracer makes it possible, for instance, to find out the proportion by which each source is used by the plant, and so find out the efficiency of the fertilizer use, the strength of the soil as a N source, and several other agronomic features of the system. An important concept is the amount of nitrogen in the plant derived from the fertilizer, termed Ndff, calculated according to:

$$\%\,\text{Ndff} = \frac{\text{atom }\%\ ^{15}\text{N excess in plant material}}{\text{atom }\%\ ^{15}\text{N excess in fertilizer}}100 \qquad [3]$$

If in the above example the analysis of the plant results in an abundance of 7.855%, or in an atom excess of 7.490%, the Ndff will be 65%, meaning that, of all nitrogen in the plant, 65% comes from the fertilizer and 35% from the soil.

Another good example of the use of stable isotopes as tracers is the evaluation of atmospheric nitrogen fixation by legumes. Legume crops have the ability to fix atmospheric N_2, which has also the natural abundance of 0.365%, an agronomically very important fact because this means they have a lower demand for fertilizer N. When legumes are grown together with other nonfixing crops, under the same conditions and fertilized with ^{15}N-enriched material, the legume has three N sources: soil, fertilizer, and atmosphere, while the nonfixing crop has only the soil and the fertilizer. Therefore, at a given time, the enrichment of the legume is lower in comparison with the nonfixing plant. By difference it is possible to measure the amount of nitrogen derived from the atmosphere, Ndfa, absorbed by the legume:

$$\%\,\text{Ndfa} = \left[1 - \frac{\text{atom }\%\ ^{15}\text{N excess in legume}}{\text{atom }\%\ ^{15}\text{N excess in nonfixing crop}}\right]100 \qquad [4]$$

Another way of using stable isotopes as tracers is through natural variations in their abundance. As already mentioned, dynamic processes lead to isotopic discrimination, and, since these processes are active all over the environment, small but measurable variations in isotopic abundance occur in nature. Plants discriminate ^{15}N from ^{14}N, ^{13}C from ^{12}C. Processes occurring in the soil can also discriminate these isotopes and those of water, i.e., H and O. The enormous cycling of water due to the processes of evaporation, condensation, sublimation, exchange, and others induces a large range in natural water isotopic compositions. The water in an evaporation tank becomes increasingly enriched in the heavier hydrogen and oxygen isotopes, which are left behind at a slightly higher rate than the lighter isotopes.

Since the variations in abundance are rather small, a more sensitive measure is used, which is the isotopic ratio given in delta (δ) values. For the ^{13}C/^{12}C ratio, the delta value of a given sample is expressed by:

$$\delta\ ^{13}\text{C}\% = \left(\frac{^{13}\text{C}/^{12}\text{C in sample}}{^{13}\text{C}/^{12}\text{C in standard}} - 1\right)1000 \qquad [5]$$

Convenient and reproducible standards are internationally set and are available for calibration purposes.

The most commonly used isotopic ratios are ^{13}C/^{12}C, ^{15}N/^{14}N, ^{2}H/^{1}H, and ^{18}O/^{16}O (Table 2). They are used in a great variety of environmental studies. Plants, for instance, discriminate isotopes in

Table 2 Principal stable tracer isotopes used for soil, plant, and environmental studies

Element	*Isotope*	*Abundance*
Carbon	^{12}C	99.985
	^{13}C	0.015
	^{14}C	Traces
Hydrogen	^{1}H	98.89
	^{2}H	1.11
	^{3}H	Traces
Nitrogen	^{14}N	99.635
	^{15}N	0.365
Oxygen	^{16}O	99.759
	^{17}O	0.037
	^{18}O	0.204
Sulfur	^{32}S	95.0
	^{33}S	0.76
	^{34}S	4.22
	^{36}S	0.014

the photosynthesis process. Thus C_3 (Calvin cycle) species have $\delta^{13}C\%$ values of −25 to −30, while C_4 (Hatch and Slack pathway) species, of −11 to −14. This makes it possible to study carbon cycling in different ecosystems, as is the case in comparisons of natural forests C_3 with introduced pastures C_4. Differences in the pasture intake by cattle can also be studied by analyzing the isotopic ratios in their feces.

Hydrologic studies use the $^{18}O/^{16}O$ ratio in many different ways. This ratio is different for sea water (taken as standard), fresh water (to the extend that different water bodies can present consistent differences), rainfall, water, soil water, etc.

Use of Radiation in Soil–Plant Investigations

Gamma Attenuation

Gamma rays are electromagnetic waves of the same nature as light, having much shorter wavelength, and consequently much higher frequency and energy. They are emitted by unstable radioactive nuclei and penetrate into matter, opening the possibility to explore the inside of materials. A gamma-ray beam is attenuated during the travel through matter, and this process is used to characterize the matter in question. Several isotopes are used as gamma sources, but the most frequently employed ones are ^{60}Co, ^{137}Cs, and ^{241}Am. ^{60}Co has two gamma rays of high energy: 1.17 MeV and 1.33 MeV; ^{137}Cs has one gamma of 661 KeV, and ^{241}Am also one gamma of low energy, 60 KeV. Beams of these gamma rays are used to study soil physical properties such as bulk density (*d*; kilograms per cubic meter) and water content (θ; cubic meters per cubic meter). The gamma attenuation follows the Beer law and, applied to a moist soil sample, is as follows (see Figure 1):

$$I = I_0 \exp[-x(\mu_s d + \mu_w \theta)] \qquad [6]$$

The incident I_0 and the emergent I beams can easily be measured with scintillation detectors coupled to electronic equipment that have an output in cps or cpm. The mass attenuation coefficients μ_s of the soil material and μ_w (square meters per kilogram) of water have to be known, but are also easily measured experimentally. The sample thickness x (meters) has also to be measured, and in this way d can be calculated when the soil water content θ is known, or θ is calculated when d is known. Using a double-energy gamma-ray beam, both d and θ can be measured simultaneously.

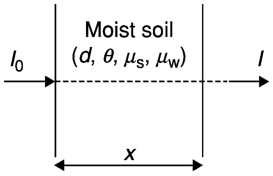


Figure 1

The great advantage of this technique is that it is nondestructive, the gamma-ray beam passes through the sample with no interference. Therefore, it is possible to make repeated measurements at the very same point, like following soil-moisture changes with time in a pot with a growing plant. Since the sample thickness x has to be known, the samples have to be regular-shaped or placed in containers. Although being developed for the laboratory, a few instruments use this technique in the field; they are gamma probes, for soil surface or greater depths, when introduced in the soil using aluminum access tubes.

Tomography

Computed tomography (CT) is a fairly new technique in soil science. It is basically the same as gamma- or X-ray attenuation, but the result is a picture of a cross-section of the sample showing a bidimensional distribution of d and/or θ. While the sample rotates along the plane of the desired cross-section, a large number of I_0 and I measurements are made without the need to know each value of x. A computer program processes the data in order to obtain the image. Only small, first-generation tomographs are used in soil research, but, in the near future, third-generation instruments will give instantaneous pictures of d and θ distributions in soil samples.

Neutron Moderation

This physical process has been used throughout the second half of the twentieth century to measure successfully soil water contents in the field. When a source of fast neutrons (high energy, 2–10 MeV) is placed into a soil, these neutrons become moderated or slowed down by collisions with the atoms of the surroundings, i.e., they lose energy down to the order of fractions of electronvolts. The most suitable neutron source used today is of ^{241}Am. The moderation process is proportional to the H content of the sample, and since H is part of soil water, the number of slow neutrons is proportional to soil-water content. A detector, sensitive only to slow neutrons, gives us their count rate in cps or cpm. An adequate calibration between the count rate and θ, for samples of known θ, makes it possible to estimate soil-water contents in other situations.

Neutron probes are commercially available instruments to measure soil-water content in the field. There are surface and depth neutron probes, and also some combined with gamma probes, to measure simultaneously d and θ. They are widely used to measure soil-water availability to crops, irrigation-water needs, and water balances of crops and natural ecosystems.

Sterilization and Mutation Breeding

At very high intensities, gamma and other types of radiation have a sterilizing effect and can be used to sterilize materials. Soils and other laboratory materials are sterilized by irradiation. For this, mostly gamma but also neutron sources of high intensity are confined in special irradiation chambers. The dose to be applied depends on the distance between sample and source, strength, type, and energy of the radiation, and exposure time. In biological materials, radiation has also the property of inducing mutations. In many situations plant breeders induce mutations in plants, to increase the probability of obtaining beneficial genetic changes in agronomically important crops.

Other Nuclear Techniques

Activation Analysis

If there are some stable trace elements of interest in a sample that are difficult to measure or which cannot be easily extracted for detection, the activation analysis can be useful. It is a method of quantitative analysis in which the sample is exposed to a flux of bombarding particles, usually slow neutrons in a reactor or fast neutrons in a neutron generator, in order to provoke nuclear reactions with the elements of interest within the sample. The stable isotopes to be determined are thereby converted into radioactive isotopes that can be assayed by usual detection techniques. This technique has been used with success in a wide range of studies, for example in determining trace elements in soils, fertilizers, and plants, and determining the provenance of archeological specimens.

X-Ray Fluorescence

The principle of X-ray fluorescence spectrography is that characteristic secondary X-rays are produced when cathode rays of high energy interact with a target of a particular element. This means that every element emits its own secondary X-ray spectrum when bombarded by primary high-energy X-rays. In the case of complex samples such as soil, it is necessary to separate and identify each element of interest, recognizing their specific secondary X-rays. The sample of the material to be analyzed is irradiated with a beam of high-energy X-rays, and the fluorescent X-rays produced in the material are directed through a crystal, following the Bragg law, and counted by a specific counter. Spectral analysis is made automatically by computed multichannel systems. The most sensitive elements suitable for this analysis lie between ^{21}Sc and ^{56}Ba, being ideal for ^{30}Zn, ^{38}Sr, and ^{42}Mo.

Autoradiography

Ionizing particles emitted by radioisotopes have the same photochemical effect on photographic emulsions as do light rays. If the radioisotope of interest emits ionizing radiation of adequate energy, i.e., with a relatively short range to induce a clear image on the photographic emulsion, an autoradiograph can be made. The plant specimen containing the radioisotope, such as root, leaf, seedling, is closely applied to a photographic film and left for an appropriate time. An image is obtained after development of the film, which permits ready location and visualization of the distribution and concentration of the tracer.

C-14 dating The radioisotope ^{14}C has a half-life of 5720 years and, since it is continuously produced by sunlight-induced nuclear reactions that take place in the atmosphere, its concentration is fairly constant in atmospheric CO_2. Metabolic processes in living organisms also ensure constant C and ^{14}C concentrations in their bodies; however, after death, the carbon exchange with external sources stops, and the ^{14}C concentration decreases gradually according to the exponential radioactive disintegration law. Therefore, the ^{14}C concentration of paleobiological materials will be much lower and proportional to their age. This technique is called carbon dating, and has been used successfully in many situations, including determinations of organic carbon turnover rates in soil.

See also: **Nitrogen in Soils:** Cycle; Plant Uptake

Further Reading

Bacchi OOS, Reichardt K, and Calvache M (2002) *Neutron and Gamma Probes and Their Use in Agronomy.* IAEA Tecdoc. Vienna, Austria: International Atomic Energy Agency.

Gusmán MEJ (1989) *Nucleonica Basica*. Bogotá, Colombia: Instituto de Asuntos Nucleares.

Hardarson G (1990) *Use of Nuclear Techniques in Studies of Soil–Plant Relationships*. Training Course Series No. 2. Vienna, Austria: International Atomic Energy Agency.

L'Annunziata MF (1987) *Radionuclide Tracers*. San Diego, CA: Academic Press.

Slater RJ (1990) *Radioisotopes in Biology: A Practical Approach*. New York: Oxford University Press.

Vose PB (1980) *Introduction to Nuclear Techniques in Agronomy and Plant Biology.* Oxford, UK: Pergamon Press.

ISOTROPY AND ANISOTROPY

T-C J Yeh and P Wierenga, University of Arizona, Tucson, AZ, USA
R Khaleel, Fluor Federal Services, Richland, WA, USA
R J Glass, Sandia National Laboratory, Albuquerque, NM, USA

Introduction

The directional behavior (isotropy or anisotropy) of soil hydraulic conductivity is of primary importance in analyzing rates and directions of water flow and solute transport in the vadose zone. While the hydraulic conductivity of a soil may vary from one location to another (heterogeneity), it can also vary from one direction to another at the same location (anisotropy). The directional behavior in the hydraulic conductivity of a soil governs many hydrologic processes, such as infiltration, recharge, evaporation, and runoff, as well as solute migration. Field soils with a large hydraulic-conductivity anisotropy may result in extensive lateral spreading of fluid solutes, and contaminants in the vadose zone. Such spreading can potentially impact groundwater recharge, irrigation scheduling, and plant root development and microbial activity.

Darcy Law

The Darcy law is the fundamental principle underlying the study of water flow in porous media. It states that the specific discharge of water in a porous medium is linearly proportional to the hydraulic head gradient and is in the direction of decreasing hydraulic head. The constant of proportionality is the hydraulic conductivity. A general form of Darcy law for three-dimensional flow through unsaturated media can be written as:

$$q_x = -K(h)\frac{\partial \phi}{\partial x}, \quad q_y = -K(h)\frac{\partial \phi}{\partial y}, \quad q_z = -K(h)\frac{\partial \phi}{\partial z} \qquad [1]$$

In eqn [1], q is the specific discharge (LT^{-1}) and subscripts x, y, and z denote directions. The hydraulic head, $\phi = h + z$, has a unit of length (L). The pressure head is h (L), which is positive for saturated media and is negative for unsaturated media, and z is the elevation head (L), which is positive upward. The hydraulic conductivity (LT^{-1}) is denoted by $K(h)$, which is the saturated hydraulic conductivity when the medium is fully saturated and is the unsaturated hydraulic conductivity when the media is partially saturated. The unsaturated hydraulic conductivity is a function of the pressure head, decreasing from the saturated hydraulic conductivity with increasing negative pressure head, or as the medium becomes less saturated.

At the pore scale, water flow in porous media takes place through a complex network of interconnected pores or openings. To describe such an intricate network in any exact mathematical manner is practically impossible. Consequently, the Darcy law considers only the average flow behavior over a certain volume of porous media, which must be greater than several pores. The volume over which the flow is averaged is defined as a 'control volume' (CV). Using this CV approach, the Darcy law essentially bypasses both the microscopic level, at which what happens to each fluid particle is considered, and the pore-scale level, at which the flow pattern within each pore and between pores is considered. Analysis of flow in porous media, based on the Darcy law, thus moves to the macroscopic level, at which only average phenomena over the control volume are considered. The property defined at a point in the mathematical models therefore represents an average property over a CV and, thus, the property at every point in space varies smoothly such that the differential calculus applies. The medium and flow are subsequently being considered as the darcian continua. This continuum concept is parallel to the continuum hypothesis in fluid mechanics and other branches of sciences.

Control Volume, Representative Control Volume, Homogeneity, and Heterogeneity

Using the CV concept, porous media can be classified as homogeneous or heterogeneous in terms of hydraulic conductivity. If the hydraulic conductivity defined over a CV is constant throughout the entire porous medium in spite of the location of the CV, the medium is said to be homogeneous. Mathematically, homogeneity means that the hydraulic conductivity of the medium does not depend on the location. Under this condition, the CV is a representative elementary control volume (REV), implying that the hydraulic conductivity measured over the CV in any part of the medium is representative of the entire medium. This definition of REV mandates the size of REV to be much smaller than the entire medium such that the hydraulic conductivity defined over the REV is constant regardless of the location of the REV.

On the other hand, if the hydraulic conductivity defined over a CV varies from location to location, the medium is then characterized as being heterogeneous. Then, the REV does not exist for this medium. Mathematically, 'heterogeneity' implies that the hydraulic conductivity of a medium is a function of the location.

Isotropy and Anisotropy

The hydraulic conductivity over a CV at any point in a medium can be either isotropic or anisotropic. The hydraulic conductivity isotropy implies that hydraulic conductivity values of a medium are independent of the direction. Mathematically, an isotropic hydraulic conductivity is a scalar, a quantity with magnitude but no direction. Conversely, if the hydraulic conductivity over a CV differs for different directions, the hydraulic conductivity of the CV is said to be anisotropic. For example, the hydraulic conductivity of the CV measured in the horizontal direction may be greater or smaller than in the vertical. In general, a porous medium may be homogeneous and nevertheless anisotropic, or it may be heterogeneous and yet isotropic at each location. Under most field conditions, soils exhibit both heterogeneity and anisotropy.

To explain the causes of anisotropy in hydraulic conductivity, steady-state, saturated flow through a series of inclined layers of homogeneous porous media can be considered (Figure 1a). Each layer has a different hydraulic conductivity value, but the hydraulic conductivity of each layer is isotropic. If the upper and the lower boundary of the media are extended to infinity, the hydraulic head on the left-hand side of the medium is greater than the hydraulic head on the right-hand side, and they are maintained constant. Analysis of flow through each individual layer based on the Darcy law under the above specified conditions reveals that the specific discharge vector follows a zigzag path (solid vectors in Figure 1a) from the left to the right boundary.

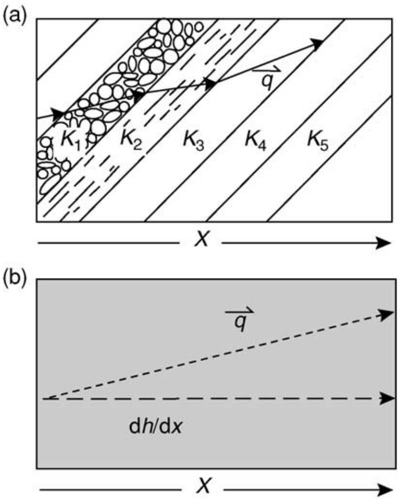


Figure 1 Schematics of flow through a layered soil (a), where K_1 is the hydraulic conductivity of each layer, $\vec{q}$ is the specific discharge vector within each layer, and an equivalent homogeneous and anisotropic soil (b) where d*h*/d*x* is the average hydraulic gradient, and $\vec{q}$ is the average specific discharge vector.

The series of layers of homogeneous porous media can be visualized as an equivalent homogeneous medium at a macroscopic level (Figure 1b) – homogenization – where the average (macroscopic) hydraulic gradient is uniform in the *x* direction (long dashed line). The average (macroscopic) specific discharge vector (short dashed line in Figure 1b) is no longer in the same direction as the average gradient. To produce such conditions, the hydraulic conductivity in the Darcy law for the three-dimensional homogeneous media thus becomes a hydraulic conductivity tensor with nine components. If a coordinate system is chosen in which the *x*-axis is perpendicular to the layers and the *y*- and *z*-axis are parallel to the layers, the number of components of the tensor can be reduced to three. The Darcy law in three dimensions for the equivalent homogeneous medium is then written as:

$$q_\mathrm{x} = -K_\mathrm{xx}(h)\frac{\partial \phi}{\partial x}, \quad q_\mathrm{y} = -K_\mathrm{yy}(h)\frac{\partial \phi}{\partial y},$$
$$q_\mathrm{z} = -K_\mathrm{zz}(h)\frac{\partial \phi}{\partial z} \qquad [2]$$

In eqn [2], the hydraulic conductivity is the macroscopic hydraulic conductivity – no longer the same as the local-scale hydraulic conductivity of each layer. In addition, the macroscopic hydraulic conductivity values, $K_\mathrm{xx}(h)$, $K_\mathrm{yy}(h)$, and $K_\mathrm{zz}(h)$, in *x*, *y*, and *z* directions, respectively, differ, depending on the direction. In general, the macroscopic hydraulic conductivity in the direction parallel to layers is greater than the hydraulic conductivity in the direction perpendicular to the layers. That is to say, the macroscopic hydraulic conductivity for the equivalent homogeneous medium for Figure 1a is not isotropic but anisotropic. Based on the preceding discussion, the macroscopic hydraulic conductivity anisotropy of the equivalent homogeneous medium is a result of several factors. They are namely variation in *K*s of layers (local-scale heterogeneity), layers (structures), and most importantly, the homogenization (averaging) of the layered medium as an equivalent homogeneous medium.

Scales and Anisotropy

Anisotropy in the hydraulic conductivity varies with observation scale (i.e., the volume over which

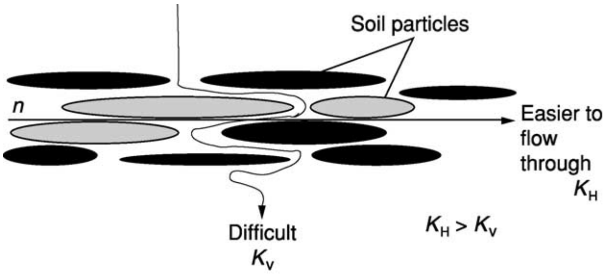


Figure 2 Pore-scale hydraulic-conductivity anisotropy, where K_v is the effective hydraulic conductivity in the vertical direction and K_H is the effective hydraulic conductivity in the horizontal direction.

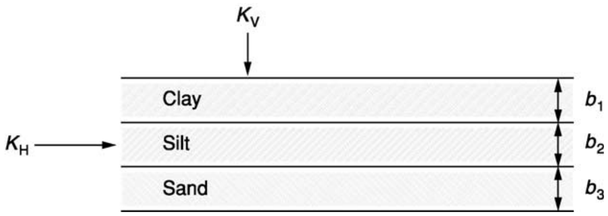


Figure 3 Field-scale hydraulic-conductivity anisotropy, caused by strata of different materials. K_v is the effective hydraulic conductivity in the vertical direction and K_H is the effective hydraulic conductivity in the horizontal direction. The thickness of each layer is denoted by b_1.

homogenization takes effect), as well as the scale of heterogeneity within the observation scale. Pore-scale hydraulic-conductivity anisotropy arises because the macroscopic hydraulic conductivity of a soil is determined over a certain volume of soil (e.g., a soil core). Within the soil volume, depositional processes cause flat particles (minerals) to orient themselves with the longest dimension parallel to the plane on which they settle. This produces flow channels parallel to the bedding plane, which allows fluid flow with little resistance. Fluid flow in the direction perpendicular to the flat surface of particles, however, must detour and take more tortuous and longer paths than for flow parallel to the bedding plane. Therefore, under the same hydraulic gradient, more flow can occur through the soil core if the gradient is parallel to the bedding plane than for perpendicular to the bedding. The hydraulic conductivity in the direction parallel to the bedding (K_H in Figure 2) is thus greater than in the direction perpendicular to the bedding (K_v in Figure 2). The soil core thus possesses a pore-scale anisotropy in hydraulic conductivity.

Field-scale hydraulic-conductivity anisotropy arises because when the hydraulic conductivity is determined in a field situation, the Darcy law assumes that there is homogeneity of the medium over a large CV. In essence, the soil is homogenized in a large CV that probably includes numerous large-scale structural heterogeneities such as stratification, cross-bedding, and clay lenses. For example, a stratified medium might be comprised of layers of clay, silt, and sand (Figure 3). Although within each layer the local-scale hydraulic conductivity may be considered homogeneous and isotropic, the macroscopic hydraulic conductivity for the equivalent homogeneous medium, an average over the conductivity values of the three layers, becomes anisotropic. In saturated media, the macroscopic hydraulic conductivity of the equivalent homogeneous medium in the direction parallel to bedding is given by a weighted arithmetic mean of hydraulic conductivity values of the layers:

$$K_H = \sum_{i=1}^{n} b_i K_i \Big/ \sum_{i=1}^{n} b_i \qquad [3]$$

In eqn [3], K_H is the macroscopic hydraulic conductivity in the direction parallel to bedding for the equivalent homogeneous medium; b_i and K_i are the thickness and the local-scale hydraulic conductivity of the ith layer, respectively; the number of layers is denoted by n. On the other hand, the macroscopic hydraulic conductivity in the direction perpendicular to bedding, K_v, for the equivalent homogeneous medium is given by a weighted harmonic mean of hydraulic conductivity values of the layers:

$$K_v = \sum_{i=1}^{n} b_i \Big/ \sum_{i=1}^{n} (b_i / K_i) \qquad [4]$$

For soil and geologic formations, where heterogeneity is not perfectly stratified, and the local-scale hydraulic conductivity exhibits a complex variation, stochastic methods have been used to derive the macroscopic hydraulic-conductivity anisotropy of the equivalent homogeneous medium. Based on stochastic analysis, the macroscopic hydraulic-conductivity anisotropy depends on several physical properties of the formations. These include the variance of local-scale hydraulic conductivity (the variation in hydraulic conductivity due to heterogeneity) and the correlation scale in different directions of the local-scale hydraulic conductivity (the average dimensions of the heterogeneity). The hydraulic-conductivity anisotropy of a fully saturated medium is therefore considered as an intrinsic property of the medium. The field-scale hydraulic-conductivity anisotropy is generally more significant than the pore-scale anisotropy for most geologic media.

Pressure Head- or Moisture-Dependent Anisotropy

For flow through unsaturated media, the unsaturated hydraulic-conductivity function can also be either isotropic or anisotropic. In contrast to the isotropy and anisotropy in saturated hydraulic conductivity, the isotropy in unsaturated hydraulic conductivity means that the unsaturated hydraulic-conductivity function is the same in all directions. The unsaturated hydraulic conductivity is anisotropic otherwise.

Few studies in the past have investigated the unsaturated hydraulic-conductivity anisotropy, especially at the pore scale. The anisotropy in unsaturated hydraulic conductivity nevertheless has often been considered as being an intrinsic property of porous media, the same as the anisotropy in saturated hydraulic conductivity. Therefore, the unsaturated hydraulic-conductivity anisotropy has been treated by scaling the unsaturated hydraulic conductivity in different directions with the anisotropy of the saturated hydraulic-conductivity. The anisotropy of the unsaturated hydraulic conductivity thus remains constant over the full range of saturation or pressure head – a constant anisotropy.

Only recently, the anisotropy in the unsaturated hydraulic conductivity at field scales has been explored. A direct averaging approach has been used as a means to approximate the anisotropy. Following the approach for saturated flow, all local-scale unsaturated hydraulic-conductivity values at a given pressure head are arithmetically averaged. This average value is then used to represent the macroscopic unsaturated hydraulic conductivity for flow parallel to bedding at that pressure head:

$$K_{\mathrm{H}}(h) = \sum_{\mathrm{i}=1}^{\mathrm{n}} b_{\mathrm{i}} K_{\mathrm{i}}(h) \Big/ \sum_{\mathrm{i}=1}^{\mathrm{n}} b_{\mathrm{i}} \qquad [5]$$

In eqn [5], $K_{\mathrm{H}}(h)$ is the macroscopic unsaturated hydraulic conductivity in the direction parallel to bedding at a given pressure head, h. The parameters b_{i} and K_{i} are the thickness and the local-scale hydraulic conductivity of the ith layer, respectively; the number of layers is denoted by n. On the other hand, the macroscopic unsaturated hydraulic conductivity in the direction perpendicular to bedding, K_{v}, for the equivalent homogeneous medium at a given pressure head is given by:

$$K_{\mathrm{v}}(h) = \sum_{\mathrm{i}=1}^{\mathrm{n}} b_{\mathrm{i}} \Big/ \sum_{\mathrm{i}=1}^{\mathrm{n}} [b_{\mathrm{i}}/K_{\mathrm{i}}(h)] \qquad [6]$$

That is, all local-scale, unsaturated hydraulic conductivity values at a given pressure head are harmonically weighted. This average value is then used to represent the macroscopic unsaturated hydraulic conductivity for flow perpendicular to bedding at that pressure head. The approximations by eqns [5] and [6], however, follow the saturated flow case and do not consider the nonlinear pressure head distribution during flow in an unsaturated medium. Consequently, anisotropy based on eqns [5] and [6] can be unrealistically large.

The macroscopic anisotropy for unsaturated hydraulic conductivity at field scales has also been derived by using a stochastic approach, which considers the nonlinear pressure-head distribution in flow processes. The stochastic formulas relate the macroscopic unsaturated hydraulic conductivity to spatial statistics of parameters of the local-scale unsaturated hydraulic conductivity. For example, if the local-scale, unsaturated hydraulic conductivity follows an exponential function:

$$K(h) = K_{\mathrm{s}} \exp(\beta h) \qquad [7]$$

where K_{s} is the saturated hydraulic conductivity and β is the pore-size distribution factor (L^{-1}), the macroscopic unsaturated hydraulic conductivity for flow parallel to bedding, K_{H}, is given as:

$$K_{\mathrm{H}}(h) = \exp\left[F + \frac{\sigma_{\mathrm{f}}^2}{2(1+B\lambda_{\mathrm{z}})} - \left(B - \frac{(2\lambda_{\mathrm{z}} - |h|)}{2(1+B\lambda_{\mathrm{z}})}\sigma_\beta^2\right)|h|\right] \qquad [8]$$

The macroscopic unsaturated hydraulic conductivity for flow perpendicular to bedding, K_{v}, is then expressed as:

$$K_{\mathrm{v}}(h) = \exp\left[F - \frac{\sigma_{\mathrm{f}}^2}{2(1+B\lambda_{\mathrm{z}})} - \left(B - \frac{(2\lambda_{\mathrm{z}} + |h|)}{2(1+B\lambda_{\mathrm{z}})}\sigma_\beta^2\right)|h|\right] \qquad [9]$$

In eqns [8] and [9], F is the mean value of $\ln K_{\mathrm{s}}$ (the natural logarithm of K_{s}), B is the mean value of β, σ_{f}^2 is the variance of $\ln K_{\mathrm{s}}$, σ_β^2 is the variance of β, and h is the mean pressure head. The formulation in eqns [8] and [9] assumes that the flow domain is perfectly stratified (λ_{z} is the average vertical thickness of the stratification (L), or the vertical correlation scale) and $\ln K_{\mathrm{s}}$ and β are uncorrelated. A ratio of eqn [8] to eqn [9] gives the anisotropy of the macroscopic-level, unsaturated hydraulic conductivity:

$$\frac{K_{\mathrm{H}}(h)}{K_{\mathrm{v}}(h)} = \exp\left[\frac{\sigma_{\mathrm{f}}^2 + \sigma_\beta^2 h^2}{(1+B\lambda_{\mathrm{z}})}\right] \qquad [10]$$

Equation [10] suggests a pressure head- or moisture-dependent anisotropy for unsaturated media. That is, the anisotropy for unsaturated K may increase with decreasing moisture content or with increasing pressure head. It may also decrease from the ratio at saturation to a ratio of 1 (isotropy) and then increase at lower degrees of saturation if K_{s} and β are

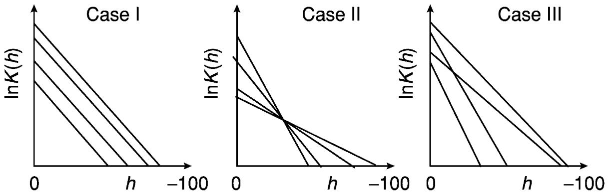


Figure 4 Local-scale, natural logarithm of unsaturated hydraulic-conductivity, ln$K(h)$, as a function of pressure head, h, in layered soils for three cases.

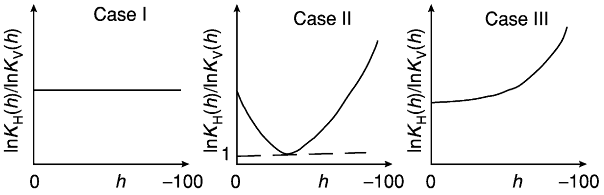


Figure 5 The corresponding anisotropy in the effective unsaturated hydraulic conductivity for the three cases shown in Figure 4. The natural logarithm of unsaturated hydraulic-conductivity in the vertical and horizontal direction is denoted by ln$K_V(h)$ and ln$K_H(h)$, respectively.

correlated. Also shown in eqn [10] is that several factors control the macroscopic unsaturated hydraulic-conductivity anisotropy. The factors are the mean, variance, and correlation scale of the saturated hydraulic conductivity and the pore-size distribution factor of the unsaturated hydraulic conductivity at local scales. More importantly, eqn [10] depicts the dependence of the anisotropy on the average pressure head or moisture content over the unsaturated media. Similar to hysteresis in moisture-retention curves, the moisture-dependent anisotropy may be subject to hysteresis effects.

While the direct averaging approach is strictly applicable to saturated media, it can be used to explain the cause of the moisture-dependent anisotropy. Consider the following three cases (Figure 4): For case I, a formation consists of many layers of porous media, in each of which the local unsaturated hydraulic conductivity has different K_s but the same β; in case II, each layer has a different K_s and β-value, but the values of K_s and β in each layer are positively correlated; case III considers each layer has a different K_s and β-value, but the K_s and β-values are independent. An arithmetic mean of all local unsaturated hydraulic conductivity values at a given pressure head yields the macroscopic unsaturated hydraulic conductivity for flow parallel to bedding at that given pressure head. In contrast, a harmonic mean is used to yield the macroscopic unsaturated hydraulic conductivity for flow perpendicular to bedding at the given pressure head.

Recognizing that a harmonic mean weighs heavily on the smallest $K(h)$, the anisotropy ratio can be plotted as a function of the pressure head for the three cases (Figure 5). Again, the anisotropy ratio is defined as the ratio of the macroscopic unsaturated hydraulic conductivity for flow parallel to bedding to the conductivity for flow perpendicular to bedding. Because the variation in $K(h)$ remains the same over the full range of the pressure head in case I, the anisotropy remains constant. In case II, the variation in $K(h)$ decreases toward a cross-over point and then increases with the pressure head. The corresponding anisotropy ratio first decreases, becomes unity (isotropic) at the crossover point, and then increases again. In case III, on the other hand, the variation in $K(h)$ always increases with the pressure and, in turn, the anisotropy grows with the pressure head.

While the moisture-dependent anisotropy has been derived for field-scale soils, similar moisture-dependent anisotropy at the pore scale has also been observed in the relative permeability of oil and gas measurements on sandstone cores. The sandstone, when dry or fully saturated, appeared to be homogeneous and isotropic. When the material was partially desaturated, however, thin and regular spaced strata became apparent. The air pemeability of the dry core was approximately twice as great parallel to the bedding planes as perpendicular to the bedding planes. Evidently, the material was quite uniform, but it was not isotropic. The effect of the anisotropy was to increase greatly the critical gas saturation and make the oil relative-permeability curve steeper when flow was across the bedding planes.

Influences of Anisotropy on Movement of Moisture Plumes in Soils

To illustrate impacts of the unsaturated hydraulic-conductivity anisotropy on water movement in unsaturated soils, the moisture content distributions after infiltration from a surface source in three different soils have been simulated. Figure 6a shows the simulated moisture-content distribution in a homogeneous soil with isotropic unsaturated hydraulic conductivity. In this homogeneous and isotropic soil, water moves predominantly in the direction of gravity. The moisture content distribution in Figure 6b corresponds to the simulated result for the same surface infiltration event in a homogeneous soil profile with a constant anisotropy in unsaturated hydraulic conductivity. Figure 6c shows the simulated moisture-content distribution in a homogeneous soil with a moisture-dependent anisotropy in unsaturated hydraulic conductivity. The moisture content distributions in Figure 6b and c show significant

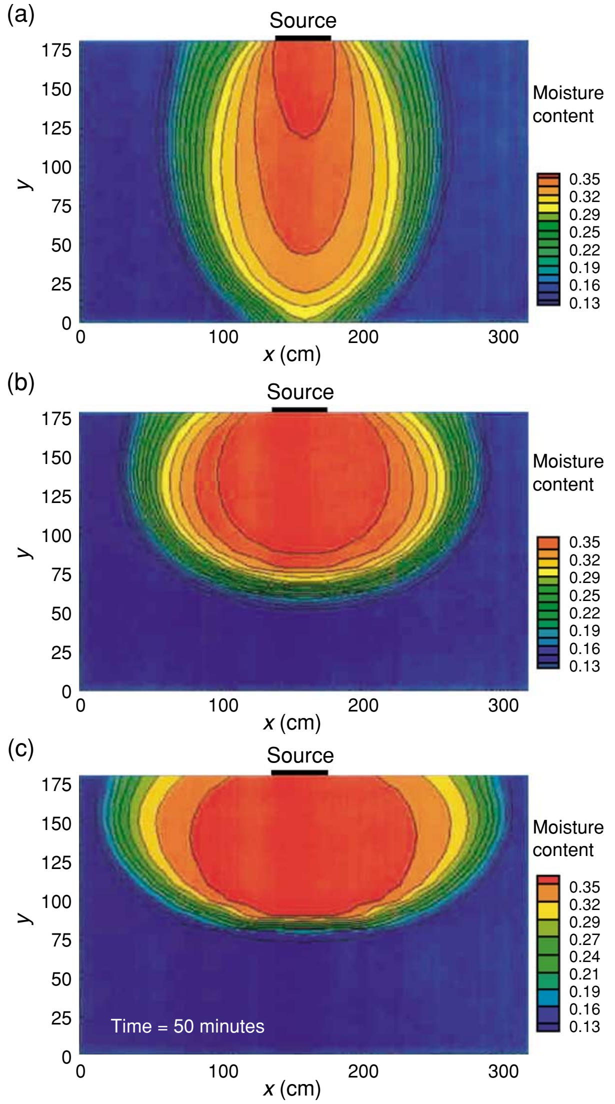


Figure 6 Simulated moisture-content contours in soil profiles after infiltration from a surface source: (a) a homogeneous and isotropic soil; (b) a homogeneous soil with a constant anisotropy; and (c) a homogeneous soil with a moisture-dependent anisotropy. The *x* and *y* denote the horizontal and vertical dimension of the soil profile, respectively.

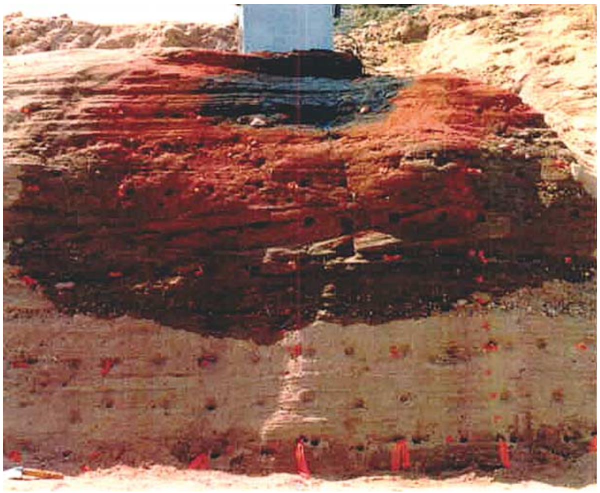


Figure 7 Moisture (dark area) and tracer (red and blue) plumes in the Rio Bravo deposits in Albuquerque, New Mexico, USA.

lateral movement of water. The anisotropy in unsaturated hydraulic conductivity in the two anisotropic soils evidently inhibits vertical movement of water but enhances its lateral spreading. However, greater lateral spreading of water in the soil profile with the moisture-dependent anisotropy is evident than in the soil with the constant anisotropy.

Effects of the anisotropy in unsaturated hydraulic conductivity have also been observed in numerous field experiments. For example, Figure 7 is a snapshot of moisture and tracer plumes in an experiment conducted in Rio Bravo alluvial deposits in Albuquerque, New Mexico. The media, according to the photo, is highly stratified, consisting of thin layers of materials of different textures. Moisture and tracer plumes spread great lateral distances and do not follow the prediction relying on a homogeneous isotropic concept – a manifestation of macroscopic anisotropy in unsaturated hydraulic conductivity. Similar phenomena have been reported in field observations of pollutant migration from a septic tank drain field in glacial outwash deposits in the Spokane Valley, Washington. At the Hanford site near southeastern Washington, extensive lateral water movement but limited vertical movement is also observed in high-level radioactive waste leakage in the unsaturated zone. Field observations in the loess-loam area north of Beer-Sheba, Israel show that rain infiltrates the soil to a limited depth with no net recharge of the groundwater – an effect of lateral flow in unsaturated soil.

More recently, detailed soil-water tracer experiments in the shallow soils of a sandy hillslope at an arid site near Socorro, New Mexico, indicate that significant horizontal downslope flow components develop despite the presence of a near-vertical downward hydraulic gradient. This suggests that, at *in situ* pressure heads, the dune sands behave as a highly anisotropic medium, although field and laboratory permeability analysis of the sands shows that the sands are nearly isotropic at complete saturation. The field and laboratory measurements have provided strong supporting evidence for the concept that, for a layered heterogeneous porous media, the macroscopic media anisotropy varies as the state (pressure, moisture, and saturation) of the media varies.

Finally, while the anisotropy is important to vadose zone processes, its effects on surface hydrology processes have also been reported. The anisotropy of unsaturated soils and the slope of the land surface can produce strong lateral flow components, which, in turn, cause moisture accumulation in concave parts of the landscape, leading to the point of saturation. The concentration of water in concave areas explains some rainfall runoff and erosion phenomena that were previously unexplained by classic concepts of infiltration.

List of Technical Nomenclature

$\boldsymbol{\beta}$	pore-size distribution parameter (L^{-1})
$\boldsymbol{\lambda}_z$	vertical correlation scale, average vertical thickness of the stratification (L)
$\boldsymbol{\sigma}_\beta^2$	variance of β (L^{-2})
$\boldsymbol{\sigma}_f^2$	variance of $\ln K_s$
$\boldsymbol{\phi}$	hydraulic head (L)
b_I	thickness of the ith layer (LT^{-1})
B	mean value of β
F	mean value of $\ln K_s$
h	pressure head (L)
$K(h)$	hydraulic conductivity (LT^{-1})
K_H	hydraulic conductivity for flow parallel to bedding (LT^{-1})
K_i	hydraulic conductivity of ith layer (LT^{-1})
K_s	saturated hydraulic conductivity (LT^{-1})
K_v	hydraulic conductivity for flow perpendicular to bedding (LT^{-1})
K_{xx}	hydraulic-conductivity tensor component in the x direction (LT^{-1})
K_{yy}	hydraulic-conductivity tensor component in the y direction (LT^{-1})
K_{zz}	hydraulic-conductivity tensor component in the z direction (LT^{-1})
$\ln K_s$	natural logarithm of K_s
n	number of layers
q_x	specific discharge in the x direction (LT^{-1})
q_y	specific discharge in the y direction (LT^{-1})
q_z	specific discharge in the z direction (LT^{-1})
z	elevation head (L)

Further Reading

Bear J (1972) *Dynamics of Fluids in Porous Media*. New York: Elsevier.

Corey AT and Rathjens CH (1956) Effects of stratification on relative permeability. *Journal of Petroleum Technology* 69.

Crosby JW, Johnstone DL, Drake CH, and Fenston RL (1971) Migration of pollutants in a glacial outwash environment. 3. *Water Resources Research* 7: 713–720.

Gelhar LW (1993) *Stochastic Subsurface Hydrology*. Englewood Cliffs, NJ: Prentice Hall.

Mantoglou A and Gelhar LW (1987) Stochastic modeling of large-scale transient unsaturated flow. *Water Resources Research* 23: 37–46.

McCord JT, Stephens DB, and Wilson JL (1991) Hysteresis and state-dependent anisotropy in modelling unsaturated hillslope hydrologic processes. *Water Resources Research* 27: 1501–1518.

Mualem Y (1984) Anisotropy of unsaturated soils. *Soil Science Society of America Journal* 48: 505–509.

Palmquist WN Jr and Johnson AI (1962) Vadose flow in layered and nonlayered materials. *Geological Survey Professional Paper* 450: 142–143.

Polmann DJ, Mclaughlin D, Luis S, Gelhar LW, and Ababou R (1991) Stochastic modeling of large-scale flow in heterogeneous unsaturated soils. *Water Resources Research* 27: 1447–1458.

Stephens DB and Heermann S (1988) Dependence of anisotropy on saturation in a stratified sand. *Water Resources Research* 24: 770–778.

Ursino N, Gimmi T, and Flühler H (2001) Combined effects of heterogeneity, anisotropy, and saturation on steady flow and transport: a laboratory sand tank experiment. *Water Resources Research* 37: 201–208.

Wildenschild D and Jensen KH (1999) Laboratory investigations of effective flow behavior in unsaturated heterogeneous sands. *Water Resources Research* 35: 17–27.

Yeh T-CJ (1998) Scale issues of heterogeneity in vadose zone hydrology. In: Sposito G (ed.) *Scale Dependence and Scale Invariance*. New York: Cambridge University Press.

Yeh T-CJ, Gelhar LW, and Gutjahr AL (1985) Stochastic analysis of unsaturated flow in heterogeneous soils. 1. Statistically isotropic media. *Water Resources Research* 21: 447–456.

Yeh T-CJ, Gelhar LW, and Gutjahr AL (1985) Stochastic analysis of unsaturated flow in heterogeneous soils. 2. Statistically anisotropic media with variable α. *Water Resources Research* 21: 457–464.

Zaslavsky D and Sinai G (1981) Subsurface hydrology. IV. Flow in sloping layered soil. *Journal of the Hydraulics Division, Proceedings of the American Society of Civil Engineers* 107: 53–64.

J

JENNY, HANS

R Amundson, University of California–Berkeley,
Berkeley, CA, USA

Hans Jenny, born 7 February 1899 in Basel, Switzerland, had a career that spanned nearly 70 years, ending only with his death on 9 January 1992. His youth was marked by a rigorous academic training, balanced by outdoor activities and experience on Swiss farms. This background led to his acceptance in the Swiss Federal Institute (ETH) in Zurich, where Jenny received a diploma in agriculture in 1922 and a DSc in colloid chemistry under Georg Wiegner in 1927 for work on ion exchange reactions. A Rockefeller Fellowship to work with Nobel Laureate Selman Waksman at Rutgers University in New Jersey, USA was followed by a position at the University of Missouri. In 1936, Jenny was appointed to a professorial position at Berkeley, which he held (as an emeritus after 1967) until his death. Jenny made significant contributions to the fields of colloid chemistry, ion exchange, and pedology. His most famous achievement is the book *Factors of Soil Formation* (1941), in which the diverse ideas of earlier scholars were combined with the abstract formalism of physical chemistry to produce a quantitative and revolutionary view of soils and ecosystems. He was the author of a biography of Eugene Hilgard, which renewed for many an insight into the significant contributions of Jenny's Berkeley predecessor. In later years, after 'retirement' in 1967, Jenny became noted for his mesmerizing lectures on soil and landscape painting and for his efforts, along with his wife Jean, on soil preservation. In 1982, at the age of 83, Jenny trekked Mount Kilimanjaro, collecting soil samples to test his hypothesis on the effect of latitude on soil formation, and subsequently performed the chemical analyses himself. Finally, he remained devoted to the science and preservation of his beloved Pygmy Forest, and the region of ancient, impoverished soils on the Ecological Staircase of Mendocino County, California. Today, Jenny's concepts on soil formation reverberate through the natural sciences. His work on soil organic matter established an observational and modeling framework that is the standard for modern soil carbon-cycling research, and his forward-looking ideas on soil esthetics and preservation are becoming topics of general research and application in the conservation arena.

Hans Jenny was born to Johann Jenny, a business accountant, who was himself the son of Johann, the owner of a chimneysweep business. Jenny's mother, Marie T. Althaus, was the daughter of a family of wood-carving factory owners in Meiringen. Jenny's first years in grammar school were in Basel. During his grammar-school period, Jenny's entire family moved to Meiringen, to live with Jenny's mother's family, because Jenny's father's business expertise was needed in the factory. Jenny remembered this period fondly. His grandfathers and uncles were artists, and they discussed and created modern art, an experience that early helped forge a lifelong interest in art and esthetics in the young Jenny. This period also introduced the young Jenny to farm life, and he spent much time on the farm adjacent to his grandparents' home (Figure 1).

Upon the family's return to Basel, his parents decided that he should attend a junior high school that focused on college preparation. As Jenny recollected:

> rigorous training began that transformed a näive adolescent into a sophisticated, somewhat smart-alecky young intellectual. . . . [but] throughout this period World War I raged at the Swiss frontiers.

Classes began early in the day, and homework kept the young students busy until after midnight. The demands on Jenny through the high-school period were traumatic. Great stress was placed on studies, particularly on rote memorization in modern languages, history, science, and mathematics. Jenny later recalled that he had a hard time memorizing words, and that because of broad-ranging extracurricular activities and tension in his parents' relationship he was just an "average student. . . and I know I had my share of feelings of frustration, and of being

Figure 1 Hans Jenny, early 1900s.

mediocre, which was considered worse than being a criminal." Jenny's 'mediocre' performance in school abruptly changed when he later entered the university.

In high school, Jenny joined the Boy Scouts, which proved to be an opportunity for him to excel in a different arena of activity and to develop leadership qualities. Jenny's particular troop emphasized an 'ascetic' approach to scouting: fasting, no display of merit badges, and no vices. Jenny was good with ropes, compasses, etc., skills which he attributed to his previous experiences on the farm. At the same time, the war raged around Switzerland, and Jenny would hear the thunder of cannons at night and see planes escaping into Basel pursued by antiaircraft fire. The wounded German and French soldiers in Basel, the families who had lost sons, and the Allied and German war propaganda left a strong impression on Jenny and his friends. Food shortages in Europe became prominent, and Jenny spent summers during his high-school years working on a farm in the Emmenthal Valley, eventually being placed in charge of horses, oxen, and plowing.

Figure 2 Georg Wiegner (left) and Hans Jenny (right), around 1925.

When his high-schooling ended in 1918, he was awarded a certificate to enter any Swiss university and chose to study agriculture. However, Jenny first spent a year working on farms in both German and French Switzerland. This period provided him with much practical experience about farming, and he entertained the notion of becoming a livestock manager following his studies at the university. The farming experience also had other effects on Jenny. The acquisition of physical skills, and the additional year of maturity, gave him a more serious perspective as a university student, and Jenny there focused his efforts on the educational opportunities that surrounded him. Many of the faculty at Zurich were internationally famous, and, as a student in agriculture, Jenny eventually took courses from Georg Wiegner in agricultural chemistry, an exposure that shifted his interests toward colloid chemistry and a postgraduate degree. Most notably, Jenny was attracted to Wiegner because he "expounded on principles, concepts, and ideas, rather than practical recipes."

In 1923, Jenny received his diploma in agriculture, and approached Wiegner about research in soil chemistry. Wiegner responded by insisting that Jenny first acquire a more solid foundation in chemistry. Wiegner was widely famous, and attracted students and postdoctoral scholars from across Europe as well as from overseas. Soon, Jenny became Wiegner's chief assistant in soil and colloid chemistry. In contrast to Wiegner's laboratory interests, Jenny began to spend weekends in the field, collecting and subsequently analyzing soil samples (Figure 2). This activity brought him into contact with Josiah Braun-Blanquet, with whom he later collaborated on the relationship of soil formation and plant development.

As a result of Wiegner's preeminence in ion exchange, Jenny was encouraged to study the individual behavior of ions during exchange in relation to their degree of hydration. Jenny found that exchange behavior could not be explained by the activities of the ions, but instead must involve complex reactions between mineral surfaces and the hydrated forms of ions. Wiegner and his students moved forward on this problem from several fronts, including conducting the exchange reactions in alcohol, in which the ions were dehydrated. Jenny recalled that the laboratory environment in Wiegner's group was interactive and dynamic, and that "doing research was an exciting cross-fertilization experience." In fact, the field of colloid chemistry at that time held a central position in chemistry and physics, a situation that made research even more exciting for the young scientists in the laboratory. Among the great advancements in that time period was Einstein's Theory of Brownian Movement, which linked molecular theory with the behavior of colloidal particles, revealing a continuity of molecules and atoms and colloidal particles that provided an overall view of the corpuscular system of nature.

Yet, on a completely different scale, Jenny noted that he had developed a reputation, within a small circle, through his interest in linking laboratory and theoretical colloidal chemistry with soils in nature. Working with Dr. Braun-Blanquet, who was an innovator in the new field of plant sociology, they explored the Swiss National Park, linking soil properties with plant distribution. In this report, Jenny drew a speculative curve relating soil organic matter to climate, relying on some of Eugene Hilgard's analyses (whose position Jenny was later to fill at Berkeley). The book, entitled *Vegetation Development and Soil Formation*, was reportedly later called a classic by R. Tüxen in the 1950s, but Jenny jokingly called it a *Jugendsünde* (an illegitimate product of youth).

Near the end of Jenny's doctoral research, Dean Mann from Cornell University, representing the Rockefeller Foundation's International Education Board, visited Wiegner's laboratory and asked Jenny if he was interested in spending a year at an American university. This was something Jenny had already considered, and he asked if he could go to Berkeley to work with Dennis Hoagland in plant nutrition. Mann discouraged this idea, suggesting that Jenny select an Atlantic-coast institution. Shortly after this, Selman Waksman from Rutgers visited the Wiegner laboratory and suggested that Jenny apply to work in his laboratory in New Jersey. Jenny's application, with Waksman's support, was approved and in 1926 Jenny left for America. The voyage itself proved to be memorable for Jenny, for he was paid to travel first class (and was therefore obligated to purchase a tuxedo), but he spent much of the time being seasick rather than enjoying the amenities and improving his English.

The short year at Rutgers was stimulating for Jenny. Waksman proposed that Jenny conduct litter-decomposition experiments, examine the colloidal character of the humus produced, and characterize the microbes involved. In later years Jenny remarked that the questions and approaches were sound, but the dismal conditions of Waksman's basement laboratory (occupants were forced to wear overboots during rain storms) and the sloppy behavior of the students made Jenny decide to switch his research to the effect of ion type on plant transpiration. This work was conducted with Professor Shive. Jenny remarked that Waksman, who was probably disappointed that Jenny switched laboratories, remained a professional and personal friend until his death, and that, although Jenny never completed his research with Waksman, he was stimulated by the ideas and discussions that Waksman's group engaged in. Jenny ultimately recognized that laboratory conditions were not always indicative of success, for he remarked: "Years later I realized that laboratory life-styles are not crucial, for it was Waksman who got a Nobel prize, not Wiegner."

The American lifestyle encountered by Jenny proved to confirm some of his European biases and debunk others. Jenny was shocked by the gum-chewing of students, and their lack of interest in international affairs, philosophy, and art. He said: "What surprised me right away was that everybody put their feet on the table or another thing, so I drew that …. [entitled] A Landscape in the USA!" (Figure 3). On the other hand, Jenny was greatly impressed by the experimental ingenuity of American chemists, and by facets of American chemistry which Europeans had ignored or neglected.

While at Rutgers, Jenny and a Danish colleague, J.H. Blom, were asked to translate submissions to the upcoming First International Congress of Soil Science, organized by Dean Jacob Lippman at Rutgers, which was to be held in Washington, DC. Jenny later remarked that he and Blom were "quite malicious" and "had fun in the translations, giving a certain rhythm as that of a hexameter or a pentameter, making fun of pompous writing." As the Congress approached, Wiegner was able to have Jenny appointed as an official Swiss delegate, which reduced the financial burden and also enabled him to

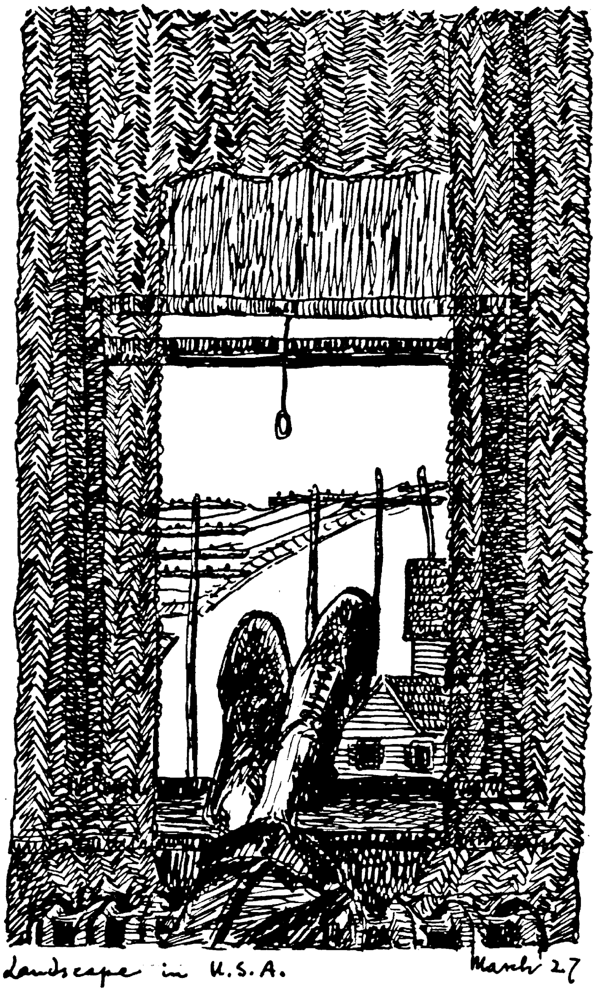


Figure 3 Hans Jenny's first sketch in the USA, 1927.

attend the important postcongress, transcontinental soils excursion.

The postcongress excursion appears to have been a remarkable trip, bringing forth both cultural and scientific integration. A select group of scientists of differing ages and backgrounds were confined to a private train, with Pullman cars, that traversed North America from the Atlantic to the Pacific. Jenny later remarked on the tensions and conflicts between the sometimes abrasive youth and the older, established leaders – as well as between scientists of differing backgrounds. Importantly, the trip, which passed through enormous ranges of climate, opened Jenny's mind further to the relationships between soil types: "The rolling plains, I fancied, must harbor the secret of mathematical soil functions. At times I could hardly sleep thinking about it." During the Congress, Robert Bradfield, from the University of Missouri, decided to go to Wiegner's laboratory for a sabbatical, and offered his laboratory and teaching position to Jenny for the year, which Jenny "accepted with great pleasure."

In the autumn of 1927, Jenny arrived in Columbia, Missouri. The physically isolated but vibrant academic setting enabled him to follow through on the ideas which were conceived on the excursion earlier that summer (Figure 4). Professor K.K. Krusekopf informed Jenny of the availability of soil C and N data from various states, and Jenny then began assembling data from across the Plains states, first arranging it by temperature (Canada to Louisiana). Jenny developed equations to describe the trends

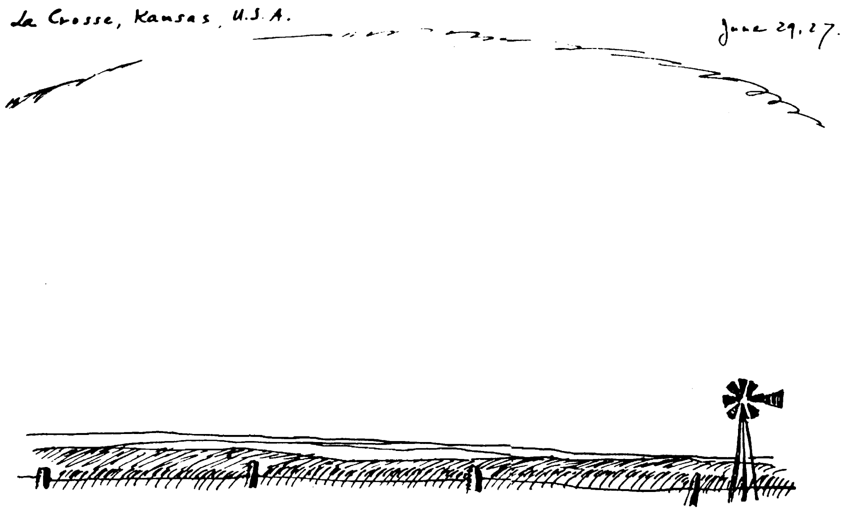


Figure 4 Jenny's sketch of Kansas made during the Transcontinental Excursion, 1927.

("primitive modeling" as he later described it) and showed it to his colleagues, who encouraged him and suggested he present the paper that November at the meetings of the American Soil Survey Association in Chicago. The paper, which was subsequently published, quantitatively explored the climatic effects on soil N, and this effort earned Jenny a research award from the American Society of Agronomy in 1931. In this and subsequent papers, data were aligned in graphs and described mathematically, and Jenny remarked:

> I enjoyed seeing field data aligned by equations and derived aesthetic pleasure from the shapes of the curves. Several pedologists, however, accused me of trying to be erudite.

In addition to his work on the climatic effects on soil N, Professor M.F. Miller suggested to Jenny that he examine the effects of cultivation on soil organic matter. A study between a native pasture soil and an adjacent cultivated field, along with data derived from the literature, produced a seminal paper and report on the loss of C and N during cultivation, and its rate. Jenny later remarked that this study added the factor of 'time' to his growing list of what would later become the famous factors of soil formation.

In addition to Jenny's pedological research at Missouri, he was greatly engaged in renewing and expanding his research on colloids and ion exchange. This interest was partially the result of large conceptual differences held by Robert Bradfield, Jenny's Missouri colleague, and Wiegner. One of the major experimental differences in Missouri was that Jenny worked with natural clays, in contrast to artificially made colloidal gels used in Zurich. A result of this work was a statistical model of ion exchange, one derived during late hours at night in his basement office during the hot and humid Missouri summer. About this period, and his parallel work in pedology and colloid chemistry, Jenny asked himself: "Am I doing the right thing… but both avenues proved interesting and beneficial to me, so why should I give up either one? I could have two sweethearts." An important, but not widely known additional area that Jenny also explored at this time was the differential behavior of K and Na during chemical weathering, work in a topic now widely recognized as geochemistry. This work, published in the Missouri Agricultural Experiment Station Bulletin was, as Jenny noted, not widely read. But Jenny sent a copy to V.M. Goldschmidt, who later cited the work in his discourses on geochemistry.

With the Great Depression still in full force, the University of Missouri was hit by financial difficulties. Young faculty were laid off, though Jenny (then an assistant professor) was spared. In 1934, faculty members were forced to take a year's leave of absence without pay, and Jenny's turn was set for 1934–35. With the assistance of his old Zurich friend Max Kleiber (then at the University of California–Davis), Jenny was able to obtain an appointment at the Citrus Research Station at Riverside, California. There, Jenny collaborated with W.P. Kelley, an arid-lands soil chemist and the codiscoverer of the crystalline nature of soil clay minerals. Jenny also, with E.R. Parker, developed a study to examine the role of tillage on water infiltration, and the time required for adverse tillage effects to disappear. While at Riverside, Kelley encouraged Jenny to visit Berkeley, where he delivered one or two lectures in the Department of Plant Nutrition. Soon, Jenny was offered the position of Associate Professor of Soil Chemistry and Morphology, which he accepted.

Upon his arrival at Berkeley, Jenny became a member of the Plant Nutrition Department, under the leadership of the famous plant nutritionist Dennis Hoagland. The Department contained numerous faculty, including Roy Overstreet. Overstreet, a soil physical chemist, was a colleague with whom Jenny was able critically to discuss his ideas and with whom he conducted joint research. Jenny repeatedly recounted the great scientific and personal pleasure that his discussions with Overstreet produced, and Jenny gives great credit to Overstreet for rigorously challenging his ideas and concepts. At the time of Jenny's arrival, the Department of Soil Technology, headed by Charles Shaw, contained people devoted to more pragmatic soil research agendas. Jenny notes that conflicts between the two groups existed, but that Jenny's interests in pedology led him to interact with Shaw, who told Jenny he (Jenny) "was going to be the new Dokuchaev." Shaw suddenly died, and Jenny and Overstreet, along with other soil researchers (soils people), then moved to a new Department of Soils.

The breadth and depth of Jenny's scientific accomplishments at Berkeley are truly remarkable. He later described his preceding years in Missouri as his '*Sturm und Drang*' (storm and stress) period because, while he was rapidly developing ideas (and publishing) in colloidal chemistry and pedology, he was haunted with insecurities ("Sometimes you have nightmares … is what you are doing all wrong?"). Berkeley provided Jenny with colleagues and collaborators who challenged him and questioned him, but encouraged him to become more confident in his emerging research.

Upon his arrival, Jenny continued to pursue his twin 'sweethearts' of colloid chemistry and pedology.

It is instructive to remember that his hiring was based on his expertise in colloids, and he immediately began teaching a rigorous and widely attended course on the topic – one attended by students from across the campus. In turn, in preparation for his courses and research, Jenny himself also audited classes on optical mineralogy, geography, and mathematics. In terms of research in colloidal chemistry/ion exchange, some of Jenny's achievements are remarkable.

First, Jenny's ion exchange expertise was applied to a problem, brought to him by Hoagland, regarding the use of ammonia gas for fertilization. Jenny demonstrated that ammonia could be applied directly to soil and that, through ion exchange, the ammonia would be held on exchange complexes until is was nitrified or bioaccumulated. Jenny wrote a memorandum to Hoagland illustrating the manner in which the gas could be injected into the soil, and the adsorption mechanisms involved. Jenny shared the memorandum with a representative of Shell Oil, who himself then patented the idea shortly thereafter. Though Jenny recognized the missed financial opportunities and prestige for the university and himself, he was pleased to see that basic scientific research in ion exchange could yield such widespread and useful applications to agriculture.

Second, following his earlier work at Missouri and before, and stimulated by his new colleagues (particularly Hoagland), Jenny and Overstreet set out conceptually and experimentally to explore the concept of 'contact exchange,' or the direct ion exchange between roots and soil colloids, in ion uptake by plants. Their work involved refined concepts of the nature of soil water (or the differing character of soil water adjacent to colloids and roots) and clever experiments with roots and gels to illustrate exchange reactions.

Additionally, in relation to contact exchange, Jenny and Overstreet (with the help of other colleagues) quickly recognized the utility of radioisotopes, then being produced by Berkeley's new Radiation Laboratory under the direction of Ernest Lawrence, in their research. Jenny recalled the intellectual excitement and potential that the Radiation Lab caused on the campus at the time:

> I remember a lecture given by J. Robert Oppenheimer where he actually drank a glass of water with radioactive sodium chloride in it, and then showed two or three minutes later that his fingers were radioactive. I guess he wouldn't do that anymore…

Jenny and Overstreet received solutions of radioactive potassium (treated by Lawrence's cyclotron) on a preferred basis, partly because both knew and talked to Lawrence, but also because the graduate student (Niels Edlefsen) who helped Lawrence build his first cyclotron (which resulted in Lawrence's Nobel Prize) later became a soil physicist, and "soils [were] known at the time to be very progressive."

While Chair of both his Department at Berkeley, and the smaller one at Davis, Jenny traveled weekly to meet his Davis colleagues. There, he discussed his model of monovalent ion exchange with Professor Lannes Davis. Jenny encouraged Davis and Overstreet to investigate the problem further, and they began to develop a thermodynamic-based model of ion exchange. Jenny encouraged and supported this research, but did not participate, because he felt the thermodynamic approach did not reveal the colloid chemical phenomena he wished to illuminate, and Jenny's own skills in mathematics and statistical mechanics were insufficient for him effectively to push his own approach further.

Jenny and his colleagues, inspired by exchange and ion uptake, focused on the nature and meaning of pH measurements in colloidal suspensions and mineral surfaces. Also, to examine better the mechanisms by which roots might interact with soil particles, electron microscopy of the 'mucigel' surrounding the root, and its reactivity, was undertaken. These and other studies too numerous to mention in this brief summary continued until Jenny's retirement.

It is probably fair to conclude, at least from his reputation today, that Jenny's greatest scientific achievement is his classic book *Factors of Soil Formation. A System of Quantitative Pedology.* It is also probably his most misunderstood achievement, a fact reflected in that few people ever cite the full name of the publication (and the second sentence is an essential component of the book), and that many recent pedological texts continue to state incorrectly the premise of the book, and its famous equation:

$$s = f(cl, o, r, p, t, \ldots)$$

where $s =$ is soil properties and cl, o, r, p, t are regional climate, potential biota, topography, parent material, and time, respectively. As Jenny once remarked, "It looks easy, but it's not."

The origins of the book appear to have begun back in Jenny's Swiss education and research, but he notes that the key to rigorously expanding the ideas, and putting them in a book, were driven by his assignment to teach pedology at Berkeley. Unlike his colloid chemistry course, in which he was an expert and part of the scientific mainstream, he entered pedology as an outsider, with somewhat novel views. He strove to develop a logical and integrative means of explaining soil geographical distributions. As many (including Jenny) have noted, his factors included those first discussed by Dokuchaev. But what most do not appreciate is the conceptual gulf between those earlier formulations and Jenny's – a fact that Jenny made clear in the very first sentence of his book: "As a science grows, its underlying concepts change, although the words remain the same." Jenny conceived of soils as a physical system and defined state factors as independent variables that define the properties of the system, bringing to pedology the formalisms of physical chemistry. The completed manuscript, submitted to McGraw-Hill, was at first rejected. But intervention by Robert Bradfield, Jenny's long-time associate from his Missouri days, succeeded in achieving its publication in 1941.

Today, *Factors* remains a vibrant and hotly debated work, because it is, from cover to cover, a presentation and illumination of a theory, in the same vein as other books of its type, such as Darwin's *Origin of Species*. As with Darwin, subsequent reaction was both negative and positive. Yet, as with *Origin of Species*, the ultimate benchmark for its success is gaged in its ability to explain and understand nature. The book *Factors* allows pedologists and other natural scientists to view the world as a natural outdoor experiment, providing them with the intellectual tools to unravel its history and its processes. Scores of papers and books have now been published based on the 'factorial' concepts, and this research, which now spans pedology, ecology, and geology, shows no sign of abatement. It is one of the crowning achievements of twentieth-century soil science.

Jenny retired in 1967 (Figure 5), though his creative activity might be viewed as having merely shifted, not abated. First, Jenny (with the help of his wife Jean) must be credited with the first rigorous analysis of the role of soils in landscape painting. He lectured widely on this for decades, and, for those fortunate enough to have attended these seminars, the emotional impact will never be forgotten. This lecture was recorded, in published form, as part of a Pontifical Academy of Sciences volume, conference proceedings which mainly focused on soil fertility, and ultimately its role in the sustainability of human populations. To summarize briefly Jenny's views on the esthetic value of soils, he wrote:

> Soil appeals to my senses. I like to dig in it and work it with my hands. I enjoy doing the soil-texture feel test with my fingers or kneading a clay soil, which is a short step from ceramics or sculpture. Soil has a pleasant smell. I like to sit on bare, sun-drenched ground and take in the fragrance of soil. As yet, neither touch nor smell sensations have been accorded aesthetic recognition, but colors delight painters, photographers, and writers, as well as you and me. In loess country, plowed fields on slopes show wide bands of attractive color gradations from dark browns to light yellows, caused by erosion of the surface soil. Warm brownish colors characterize fields and roofs in Cézanne's landscape paintings of southern France, and radiant red soils of the tropics dominate canvasses of Gauguin and Portinari. Soil profiles viewed in pits may reveal vivid color and structure patterns of layers or horizons. I have seen so many delicate shapes, forms, and colors in soil profiles that, to me, soils are beautiful.

For decades, Jenny had thought about and discussed additional factors of soil and ecosystem formation. Since his work in the tropics in the late 1940s, he had considered latitude to be a candidate as an additional soil-forming factor. To test the theory required the selection of sites (all with the same mean climates, parent materials, etc.), but at different latitudes. To accomplish this experimental design meant that sites at low latitudes must be found at high elevations in order to be comparable with higher-latitude sites. Mount Kilimanjaro, in Tanzania, offered an opportunity. So, in 1982, at the age of 83, Jenny and Dr. Jennifer Harden, of the USGS, traveled to Tanzania and hiked and sampled soils to an elevation of more than 13 000 f. Jenny measured the C and N content of the soils himself, upon their return. Though not published alone, the samples and data later formed a central part of an MS thesis by, fittingly,

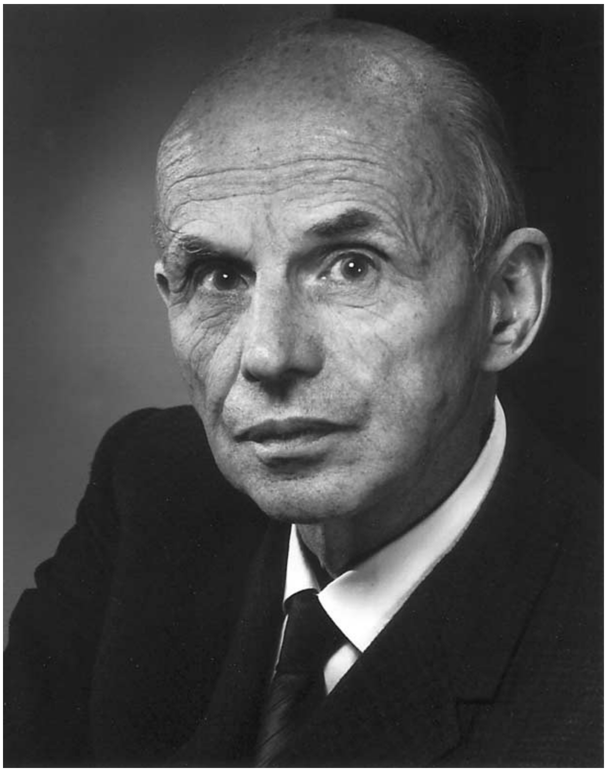


Figure 5 Hans Jenny, 1967, the year of his retirement.

a visiting student (Annelies Uebersax) from the ETH in Zurich.

For those who knew Jenny during the last decades of his life, it was apparent that his intellectual passion was fueled by a somewhat intertwined attachment to the nutrient-impoverished soils of the marine terraces of the Mendocino coast (the Ecological Staircase), and an effort to preserve them, and other rare soilscapes, for future generations. The Jenny family maintained a farm home, in nearby Comptche, which served as an overnight lodging and rustic 'think tank' for a generation of students and visitors of all types. The length of geologic time encompassed by the terraces and the revealing trends in soil chemistry enriched Jenny's views of the effect of time on soil formation, and the fate of soils on the Earth's surface. The rarity of these ecosystems, and the low regard in which they were held, led Jenny and his wife on a long journey into the workings of the politics of land preservation. Yet, despite the novel view of 'preserving soil,' they succeeded, and Jug Handle State Reserve, near Caspar, California, was created as a state-owned park devoted to leading the park visitor on a guided walk through geologic time. The visitor center prominently displays the trends of soils and ecology on the sequence of marine terraces. It is a living memorial to Jenny's insights and efforts on 'soil diversity,' many of which are being openly discussed and analyzed in the scientific literature only now. In an interview late in life, Jenny was asked:

> Does soil have a right to be protected for any reason other than that based on what is best for humans?

Jenny replied:

> Today, the idea of stewardship of land is pitted against the belief in soil exploitation for personal gain and that soil is merely an economic commodity in the marketplace ... I place soils and ecosystems, the nature museums, on a par with art museums ... colleges and temples ... Society grants human beings the right to exist, regardless of whether we are useful or not ... and the same privilege has been extended to a few endangered plant and animal species. I wish society would extend the same right to soil.

The remarkable life and genius of Hans Jenny have left an indelible stamp on soil science, forever changing the way in which soils are thought of and studied, and his ideas have diffused so widely that they are now part of the high-school curriculum. Jenny often remarked that he was honored and humbled to fill the chair of Eugene Hilgard, his predecessor at Berkeley. (*See* **Hilgard, Eugene Woldemar**). Those feelings are now appreciated by those who must follow Jenny.

See also: **Hilgard, Eugene Woldemar**

Further Reading

Amundson R and Jenny H (1997) On a state factor model of ecosystems. *Bioscience* 47: 536–543.

Amundson R, Harden J, and Singer MJ (1994) *Factors of Soil Formation. A Fiftieth Anniversary Retrospective*. SSSA Special Publication No. 33. Madison, WI: Soil Science Society of America.

Jenny H (1929) Relation of temperature to the amount of nitrogen in soils. *Soil Science* 27: 169–188.

Jenny H (1936) Simple kinetic theory of ionic exchange. I. Ions of equal valency. *Journal of Physical Chemistry* 40: 501–507.

Jenny H (1961) *E. W. Hilgard and the Birth of Modern Soil Science*. Pisa, Italy: Collana della Rivista "Agrochimia."

Jenny H (1968) The image of soil in landscape art, old and new. *Pontifical Academy of Sciences Scripta Varia* 32: 947–979.

Jenny H (1989) *Hans Jenny. Soil Scientist, Teacher, and Scholar.* Regional Oral History Office, The Bancroft Library, University of California–Berkeley, CA.

Jenny H (1994) *Factors of Soil Formation. A System of Quantitative Pedology.* New York: Dover Press. (Reprint, with Foreword by R. Amundson, of the 1941 McGraw-Hill publication.)

Jenny H and Stuart K (1984) My friend, the soil. *Journal of Soil and Water Conservation* 39: 158–161.

Jenny H, Overstreet R, and Ayers AD (1939) Contact depletion of bare roots as revealed by radioactive indicators. *Soil Science* 48: 9–24.

Jenny H, Nielsen TR, Coleman NT, and Williams DE (1950) Concerning the measurement of pH, ion activities, and membrane potentials in colloidal systems. *Science* 112: 164–167.

K

KELLOGG, CHARLES

J D Helms, USDA Natural Resources Conservation Service, Washington, DC, USA

Introduction

Charles Edwin Kellogg led one of the most distinguished careers in the history of soil science, a career marked by a dedication to assist land users through his knowledge of soils. Kellogg attended college during the formative years of the young discipline termed soil science. As a student and young professor he made landmark contributions in interpretations of soil-survey information for practical uses (Figure 1). He made significant contributions in ‘pedology,’ but his advice on using soil surveys to assist people defined his career. To accomplish that objective he redirected and refined the US Department of Agriculture (USDA) soil survey program, which he supervised from 1934 to 1971. In his later career he advised, domestically and internationally, on soils and sustainable agricultural development. He became knowledgeable about tropical soils and used that knowledge to assist developing countries.

Figure 1 Charles Edwin Kellogg, courtesy of John Tandarich.

Education and Educator

Charles Edwin Kellogg (August 2, 1902–March 9, 1980), soil scientist, was born near Ionia, Ionia County, Michigan, the son of Herbert Francis Kellogg, a farmer, and Eunice Irene Stocken. Following graduation from Palo High School, Kellogg enrolled at Michigan State College where he pursued studies in science and mathematics, and eventually specialized in the emerging field of soil science. As a student, Kellogg participated in several pioneering events in the use of soil data for planning. During the summers between 1923 and 1926, he researched and mapped soils for the land economic survey, an interdisciplinary project of the Michigan Department of Conservation, which brought together information on soils, botany, and economics, among other subjects. One objective of the survey was to help solve the problem of inefficient use of land, which vexed failing farmers and taxed the local economy and the state. From this experience and others, Kellogg learned to view soil properties in the context of their meaning to and interaction with the local community. Kellogg recognized Lee Roy Schoenmann, director of the land economic survey, as “the person who first showed me the soil,” in the dedication to his book *The Soils That Support Us*. Throughout his career, Kellogg insisted on studying soils as part of the broader natural and economic setting.

The next event stimulating Kellogg’s interest in the relationship between soils and land uses came quickly. While on a Soils Fellowship at Michigan State College (April 1926 to April 1928) sponsored by the Michigan State Highway Department, Kellogg noted failures of new cement-concrete highways in the state and made a study of the relationship between soil characteristics and appropriate highway design. In 1929, the Michigan Academy of Science, Arts and Letters published Kellogg’s article ‘Soil type as a factor in highway construction in Michigan.’ The work led to definite procedures for highway construction in accordance with the soil conditions as assessed by detailed soil surveys. The Michigan field manual of highway construction incorporated the study's recommendations. The Highway Research Board of the Public Roads Administration also issued a bulletin *Engineering Use of Agricultural Soil Maps*.

From April 1928 until December 1929, Kellogg directed soil surveys and carried out research on soil genesis for the Wisconsin Geological and Natural History Survey. While taking additional graduate courses at the University of Wisconsin, Kellogg reworked the classification of Wisconsin soils and used the results as his dissertation. Michigan State College of Agriculture and Applied Science awarded the PhD degree in 1930 for his dissertation, 'Preliminary study of the profiles of the principal soil types of Wisconsin.' The study was in the field of 'pedology,' a newly applied disciplinary term referring to descriptions and studies of soil development.

Upon completing his doctoral work, Kellogg continued to innovate in the area of interpreting soils information for utilitarian purposes. Kellogg took an assistant professorship at the North Dakota Agricultural College in January 1930, with responsibilities for the soil survey program and land classification. While at North Dakota, Kellogg advised on the design of the soil survey of McKenzie County, North Dakota, so that it could be used in land classification for rural tax assessment. From his North Dakota experience he published articles that described methods for rural land classification (*Journal of Land and Public Utility Economics*), proposed magnesium as a possible key to phosphorus problems in semiarid soils (*Journal of the American Society of Agronomy*), and described the morphology and genesis of the Solonetz soils of western North Dakota (*Journal of Soil Science*). In 1934, 'The place of soil in the biological complex' appeared in the *Scientific Monthly*. The article belonged in a genre that was to mark much of Kellogg's writing. It explained a scientific subject to the nontechnical, but educated, audience interested in public policy issues. Although Kellogg would soon depart the university, he remained an educator. During his tenure at North Dakota he inspired a number of students to make soil science a career, including Roy W. Simonson, Marlin G. Cline, A. Clifford Orvedal, and Andrew 'Andy' Aandahl.

Soil Survey Division, US Department of Agriculture

In late 1933, A. G. McCall, Chief of Soil Investigations in the Bureau of Chemistry and Soils, corresponded with Kellogg about a survey of the lands above Norris Dam in the Tennessee Valley Authority (TVA), an arrangement that did not come to fruition. Subsequently, Kellogg accepted a position in land classification in the Bureau, despite North Dakota state's bid to retain him. Roy Simonson, a Kellogg student, recalled that Henry G. Knight, Chief of the Bureau of Chemistry and Soils, visited the university to hear one of Kellogg's seminars shortly before his hiring in February 1934. It is likely that the Bureau's leadership had already identified Kellogg to succeed Curtis Marbut.

On July 1, 1934, Kellogg became Acting Chief, Division of Soil Survey, and then Chief of the Division on July 1, 1935. Kellogg manifested his strong commitment to public service through the soil survey. He reoriented the soil survey to focus on interpretations of soils data for practical uses. When the survey began under Milton Whitney's leadership, it offered the promise of helping farmers in efficient land use, perhaps an ambitious objective given the state of the science at the time. Subsequent soil-survey operations under Curtis F. Marbut highlighted viewing soil as natural bodies worthy of scientific study. Marbut took pride in establishing the soil survey as a scientific document, and consciously eschewed agricultural interpretations, leaving that task to other agricultural specialists. Kellogg believed that the soil scientists themselves should work with other technical specialties to develop the interpretations of the soil-survey.

To reach the goal of creating a more useful soil survey document required greater attention to the survey itself. The early years of Kellogg's leadership focused on tools to improve and advance the soil survey. Kellogg believed the survey needed more precision in soil definition – a belief that led to the first edition of the *Soil Survey Manual*. The manual appeared in 1937, about 3 years after Kellogg assumed leadership, and was the first major revision of field instructions in 23 years. A little earlier in 1936, he had issued a USDA publication *Development and Significance of the Great Soil Groups of the United States*, which is significant for its discussion, in qualitative terms, of the five factors of soil formation. Kellogg also commenced work on the soil classification system. The Russian concepts of soil formation, which Marbut introduced to the American audience, informed the soil classification system that first appeared in the 1938 Yearbook of Agriculture, *Soils and Men*.

Kellogg as Scientist-Administrator

Kellogg fitted the model of the scientist-administrator in the federal employ. Directly and indirectly, government-sponsored scientific investigations advance many facets of American life, through agricultural improvement, natural resources exploration and exploitation, military capabilities, and health and medicine. The government-sponsored soil survey program had far-ranging impacts on the interrelated fields of soil survey, soil classification, and soil science. Many of the methods and terminology developed in the soil survey effort formed the lexicon of soil survey and classification. The soil bodies as defined for the soil survey – the soil types and soil series – became the reference points in discussions of soils in the USA. Partly as a matter of convenience, the soil classification system developed by the soil survey, in cooperation with university-based soil scientists, became the national system. As a scientist-administrator, Kellogg was tireless in his efforts to shape the production aspects of the county soil surveys and the scientific and technical research needed to produce those surveys.

Soil Taxonomy

After World War II, Kellogg directed that a new soil classification system be developed. The staff, under Guy D. Smith's leadership, cooperated with soil scientists internationally in developing the new classification system, *Soil Taxonomy*. Smith had a keen appreciation for quantification and the scientific method and had studied with Hans Jenny, who earlier had tried to quantify the factors of soil formation. Kellogg had once remarked that there was no science without measurement. Following upon the theme of more precision in definition, division staff made greater use of laboratory analysis in definitions. Quantifiable boundaries and parameters defined the various categories in the system. *Soil Taxonomy* also sought to correct a problem that had plagued previous soil classifications: the lowest units, the soil type, did not fall logically and consistently into the upper categories. To remedy this, soil type and soil series, the lowest categories, were classified based on their measurable properties.

Kellogg hoped that the new classification system would aid transfer of empirical experience and research results internationally. If lessons in successful soil-management techniques on a particular soil could be transferred to similar soils elsewhere, it could be extremely beneficial to developing countries. For quite some time soil scientists had been keying interpretations and management recommendations to soil type. Bringing greater quantification to the designation of soil types and soil series enhanced the potential for interpretations. Despite his aspirations for *Soil Taxonomy*, Kellogg recognized that any attempt to organize knowledge about natural systems needed flexibility to adapt in the face of new discoveries. Thus he warned against letting a system become so classical that change and revision became impossible. *Soil Taxonomy* became the *lingua franca* of scientific discussion of soils in the USA and in other countries.

Soil-Geomorphology Studies

Kellogg launched a group of soil-geomorphology studies in various climatic and physiographic regions to expedite and improve soil surveys. The studies demonstrated the intricate processes of landscape and soil formation. Soil scientists could extrapolate some of the approaches and information about processes of landscape formation to other locations. Kellogg also believed that the studies would help soil scientists predict more accurately the long-range effects of soil- and water-management practices on soil stability. The studies conducted under the leadership of Robert V. Ruhe rank as seminal literature in the field of geomorphology. In addition to innovative research design and the ability to interpret the results, geomorphologic studies such as these required funding and institutional support beyond the resources of most individual earth scientists. Through Kellogg's commitment in concentrating on these projects, he and his staff profoundly influenced a field of earth science. While the studies influenced the operations of the soil survey at the time, some believe that their benefits for continuing management of the soil survey have not yet been fully exploited.

Training Soil Scientists

The last two decades of Kellogg's soil survey career marked a tremendous acceleration in the production of soil surveys. The Department of Agriculture merged the Soil Survey Division into the Soil Conservation Service (SCS) on November 19, 1952. Kellogg acquired more staff and resources to produce county soil surveys. That task presented a challenge. Kellogg's soil survey division had constituted a small corps of highly trained individuals who had contributed to the literature of soil science and established themselves in the discipline. The task before him was to create a much larger corps of field soil scientists to make soil surveys.

Kellogg prided himself on gaining governmental endorsement of the term 'soil scientist' as the official job classification, but struggled with how to ensure that this cadre of specialists had access to the most current knowledge. With the collaboration of former students and colleagues at universities, Kellogg created soil science institutes, or intensive short courses, at selected universities to give soil scientists additional training in the latest developments in the discipline. A former student, Marlin Cline, directed the first soil science institute at Cornell University. Subsequently, faculty at Iowa State University and Oregon State University hosted the soil science institute. In more recent times North Carolina State University, University of Florida, Texas A & M University, Alabama A & M University, and the University of California-Davis also hosted the institute. Kellogg also instructed the SCS soil scientists through internal memoranda and reports, which demonstrated the same care and precise, lucid writing style that characterized his articles for scholarly journals.

International Influence

After World War II, Kellogg advised farmers, scientists, and administrators internationally, tailoring his advice to fit the local educational, economic, and industrial conditions. He was nondoctrinaire in his advice. While he recognized developing countries would want to increase food production with fertilizers, he did not reflexively advocate western technology. Shifting cultivation and other indigenous methods might still be needed until fertilizer plants and agricultural infrastructure were in place. He emphasized that social and institutional barriers, in addition to scientific and technical problems, must be addressed in the developing world. His preference was not for any particular, idealized, 'fixed' pattern of land use, but for the use of science. He believed that technology, science, national needs, and markets would bring change and the soil scientists should be ready to provide the needed information for efficiency and conservation. In the post World War II period, he saw the need to increase food production, but in the context of improving other facets of the human condition worldwide. He believed that scientists should work for a balance between resource use and population. Kellogg never left the land-users out of the equation in his discussions. As western governments tried to bring agricultural and scientific advancement to the developing world, he urged that those governments also work for the freedom and security of the individual cultivator.

In time, foreign-aid specialists would recognize the faults in the post World War II technical advice provided to developing countries. Very early Kellogg saw some of the difficulties confronting the effort and the mistakes being made. He wrote that agricultural improvement was an essential component of general economic progress, not a separate activity. Kellogg noted the faulty strategy of trying to improve availability of manufactured consumer goods without also supporting farmers and agriculture. He attributed some of this to the fact that most of the planners in the developing countries hailed from the urban centers. Each country or region would need to develop its own specialists and scholars. Imperialism based on scientific expertise would be as objectionable as other types of imperialism.

Tropical Soils

In the post World War II international demand for western agricultural specialists, scientists once again had access to the tropics. Kellogg traveled, studied, wrote, and spoke on world soil resources and made tropical soils a specialty (Figure 2). On his many field

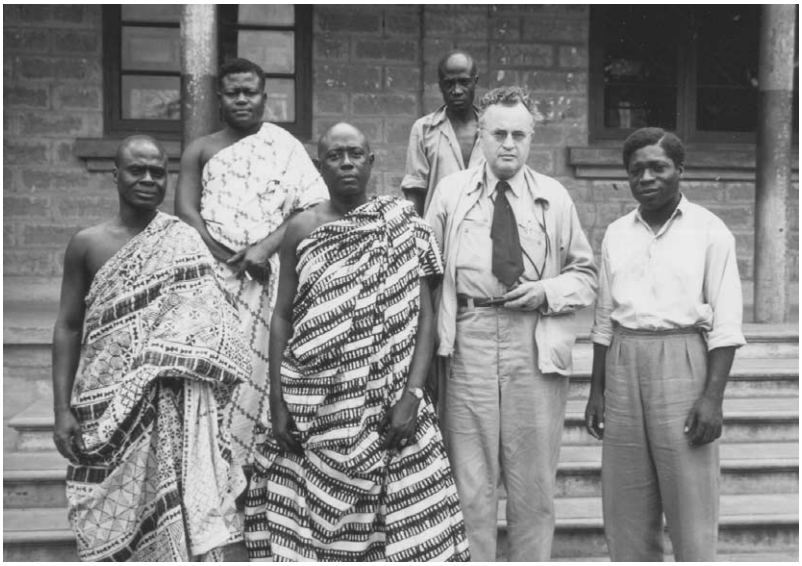


Figure 2 Charles E. Kellogg interviews farmers at Oyeko, Gold Coast (now Ghana) about cocoa production. Left of Kellogg (center) is the 'Chief Farmer' of the Gold Coast, Kwane Poku; at extreme right is Saakodee who interpreted for Kellogg. Reproduced from *African Journal K-2225* August 9, 1954. Lincoln, NE: Soil Survey Laboratory, Natural Resources Conservation Service.

trips Kellogg took copious notes, which his wife Lucille later typed into journals illustrated with properly captioned photographs. The journals remained unpublished, since Kellogg recognized that more research and analysis was needed to verify his observations, although he did incorporate some observations into his writings. In some cases, he had an opportunity for extended visits, including time for research and analysis that did lead to publication. His visit to the Belgian Congo is examined in *An Exploratory Study of Soil Groups in the Belgian Congo* (1949).

Although Kellogg might discourse on complex scientific questions, he practically always turned the discussion to questions of immediate concern to the local land-users in the developing countries. His articles emphasized that much remained to be learned about tropical soils and their management, an aspect he felt was masked by a veil of misunderstanding about tropical soils. To make the point, he employed a metaphor from art. Michelangelo supposedly said of a block of pure white marble, "A beautiful figure lies within, I have only to find it." Likewise, the agricultural scientists had to discover the combination to productivity in tropical soils. Many tropical soils had low nutrient levels, but also had great potential. As he did with domestic audiences, Kellogg emphasized that agricultural productivity was the result of interactions. Inherent soil fertility was only one element. Kellogg often cited Florida as an example of potential productivity of inherently infertile tropical soils. But he also observed that putting all of the elements in place in developing countries would take time. In the interim, western scientists should not condemn existing systems outright, but work within them to raise productivity.

Kellogg became a leading spokesperson on global soil resources. He believed that a general world soils map could have great benefit, especially in the developing world. Such a map had the potential to save time, money, and effort by making it possible to relate experience and research results from one part of the world to another. He believed there were sufficient soils worldwide that could maintain food production, if used wisely. At the same time, he was in the vanguard of the prime farmland protection movement, and he noted that development often took some of the best farmland. Use of soil information, he advised, could help guide expansion on to soils not so well suited for agriculture. He also spoke out on soil conservation. Conservation was not merely saving and maintaining soil in place or preventing degradation, but rather was the sustained and efficient use of soils that in turn contributed to improved living standards. Conservation needed to be efficient and sustaining within the larger economy.

Interactions in Agricultural Productivity

In the postwar period, the USA continued its acceleration of agricultural productivity. Kellogg was an active speaker and writer, who wove science and agriculture into public-policy recommendations. He reiterated several themes. He emphasized that the increases in agricultural productivity resulted from the interaction of individual elements of productivity – machinery, hybrid varieties, fertilizer, and water control. The individual farmer could not create this environment, favorable to productivity, alone. Many pillars supported productivity: education, extension, science and technology, industry, and markets. Interactions with particular soil types also played a role. Kellogg told soil scientists at the Seventh International Congress of Soil Science that he believed the most important principle good farmers had learned was that of the interaction among soil characteristics and among practices. This concept had particular pertinence to the developing world, where there was a great temptation to promote one item as a panacea to boost agricultural production.

Science and Policy

Two somewhat countervailing emphases are seen in Kellogg's ideas about science and public policy. He liked rigorous science and specialization, but recognized the importance of bringing in knowledge and opinions outside his discipline. So, while he saw the need for studying small soil bodies in detail, he also emphasized that interpretation of soils data be made in an interdisciplinary fashion. For example, he wrote articles about the relationship of soil use to economics and about how economics influenced potential uses. He found the distinction often made between basic and applied research as specious. He thought the USA should spend more time on basic research directed toward an objective. Unimaginative basic research was as objectionable as unimaginative applied research. While he emphasized interdisciplinary work, he warned against relying on generalists, untrained in the sciences. Rather the nation needed scientists and specialists who stood ready to work with other disciplines in an interdisciplinary fashion. His advocacy of objective-driven research influenced many, but by the end of his career, Kellogg nevertheless detected a tendency among some in soil science to study soil without any seeming advance in utilitarian knowledge.

Kellogg's published record reveals him as free from dogma, and nonideological in his advice. One is tempted to see this as a manifestation of his deep understanding of the complex interactions of soils, society, and government. According to his son, Robert Kellogg, he felt passionate about the public service role of the federal employee, but he saw that role as limited in land-use decisions. The soil survey gained new authorities in the 1960s to provide information to local governments and other entities for community planning and resource development. In the midst of this triumph he added the cautionary note. He supported the activist role of the New Deal and its capacity for giving advice and assistance in planning, providing that the decisions be made at the state and local level. The soil survey did not make recommendations. Decisions about alternative uses of tracts of land would properly be made by individuals within the regulations and restraints established by local governments under state authorities.

Kellogg educated rural America and the agricultural establishment, not just on soils but also on broad rural and agricultural policy issues. He advised planners to take account of the changing nature of rural society. As agricultural productivity grew in the USA, there was an accompanying decrease in numbers of farmers and shifts of agricultural land use. Kellogg felt strongly that people needed to distinguish between farming and agriculture. While jobs in farming had declined, other jobs in agriculture had grown. Rural land was only partly cropland or agricultural land. Rural communities had opportunities to use their lands for recreation, forestry, housing, and other activities that provided jobs. He recommended that the land-grant university system could be of assistance by establishing 'natural resources institutes,' so long as they involved local people. While Kellogg favored an active government in research and education, he often returned to the theme that local people had the responsibility to make their own land-use decisions.

See also: **Calcium and Magnesium in Soils**; **Classification Systems:** USA; **Jenny, Hans**; **Marbut, Curtis Fletcher**; **Nutrient Management**; **Shifting Cultivation**; **Tropical Soils:** Humid Tropical; **World Soil Map**

Further Reading

Effland ABW and Effland WR (1992) Soil-geomorphology studies in the US Soil Survey. *Agricultural History* 66: 189–212.

Gardner DR (1998) *The National Cooperative Soil Survey of the United States*. Historical notes no. 7. Washington, DC: Natural Resources Conservation Service, US Department of Agriculture.

Helms D, Effland ABW, and Durana P (eds) (2002) *Profiles in the History of the US Soil Survey.* Ames: Iowa State Press.

Kellogg CE (1930) *Preliminary Study of the Profiles of the Principal Soil Types of Wisconsin*. PhD dissertation. East Lansing, Michigan: Michigan State College of Agriculture and Applied Science, Soils Department.

Kellogg CE (1933) A method for the classification of rural lands for assessment in Western North Dakota. *Journal of Land and Public Utility Economics* 9: 10–15.

Kellogg CE (1934) Morphology and genesis of the Solonetz soils of Western North Dakota. *Soil Science* 38(6): 483–501.

Kellogg CE (1934) The place of the soil in the biological complex. *Scientific Monthly* 34: 46–51.

Kellogg CE (1937) *Soil Survey Manual*. Washington, DC: US Department of Agriculture Miscellaneous Publication Number 274.

Kellogg CE (1943) *The Soils That Support Us: An Introduction to the Study of Soils and Their Use by Men*. New York: Macmillan.

Kellogg CE (1948) Conflicting doctrines about soils. *Scientific Monthly* 66: 475–487.

Kellogg CE (1948) Modern soil science. *American Scientist* 36: 517–536.

Kellogg CE (1957) We seek; we learn. In: *Soils: The Yearbook of Agriculture 1957*, pp. 1–11. Washington, DC: Government Printing Office.

Kellogg CE (1963) Why a new system of soil classification? *Soil Science* 96(July): 1–5.

Kellogg CE (1968) Fit suburbia to its soils. In: *Soil, Water, and Suburbia*. Washington, DC: US Department of Agriculture, US Department of Housing and Urban Development.

Kellogg CE (1971) *Reading for Soil Scientists, Together With a Library, Revised*. Washington, DC: Soil Conservation Service, US Department of Agriculture.

Kellogg CE and Davol FD (1949) *An Exploratory Study of Soil Groups in the Belgian Congo*. Serie Scientifique No. 46. Publications de l'Institut National pour l'Etude Agronomique du Congo Belge.

Kellogg CE and Knapp DC (1966) *The College of Agriculture: Science in the Public Service*. New York: McGraw-Hill.

Kellogg CE and Orvedal AC (1969) Potentially arable soils of the world and critical measures for their use. *Advances in Agronomy* 21: 109–170.

Kellogg RL (2003) Remembering Charles Kellogg. *Soil Survey Horizons* 44: 86–92.

Simonson RW (1987) *Historical Aspects of Soil Survey and Soil Classification*. Madison, Wisconsin: Soil Science Society of America.

Simonson RW (1989) *Historical Highlights of Soil Survey and Soil Classification with Emphasis on the United States, 1899–1970*. Technical paper no. 18. Wageningen, The Netherlands: International Soil Reference and Information Centre.

Soils and Men: Yearbook of Agriculture 1938 (1938) Washington, DC: Government Printing Office.

KINETIC MODELS

P M Jardine, Oak Ridge National Laboratory,
Oak Ridge, TN, USA

Introduction

The vast complexity of subsurface chemical, physical, and microbial processes in soils, lends itself nicely to a system that is, more often than not, in a state of nonequilibrium. The mere physical structure of most soils, with its vast array of pore regimes, coupled with the transient flux of water and solutes during storm infiltration, results in a system that is often limited by physical and hydraulic nonequilibrium. Likewise, the complex mineralogical makeup of soils, coupled with water flow and media structure, often creates a system that is kinetically limited with regard to interfacial geochemical and microbiological reactions. Experimental recognition of the importance and frequency of nonequilibrium reactions and processes in soils has encouraged the reformulation of tradition equilibrium-based fate and transport models to incorporate kinetically limited chemical, physical, and microbial processes. Such mathematical models vary in complexity; encompassing single or multiple coupled processes that are useful for predicting either microscopic- or macroscopic-scale solute fate and transport behavior. The text will focus on time-dependent interactions of solutes at the solid–liquid interface in open, flowing systems. Kinetic models for soil chemical phenomena in closed systems have been well documented elsewhere.

Single-Component, Multiprocess Models

Physical Nonequilibrium

Two-region approach The occurrence of physical nonequilibrium during solute transport in soils often depends on the degree of interaction between macroscopic transport properties (i.e., water flux and preferential flow, hydrodynamic dispersion) and microscopic physical traits (i.e., diffusional mass transfer, aggregate size, length scale between pore classes), with the extent of interaction controlled by the physical structure of the media and the interconnectedness of soil water. A common mathematical representation of the time-dependent interaction of water and solutes between various pore classes within soils involves partitioning soil water into mobile macro- and mesopores (flowing region) and immobile micropores (stagnant region). The two domains are linked by an interaction term that considers the diffusional mass exchange of water and solutes from one pore class to another. Both macroscopic and microscopic descriptions of intrapore solute transfer have been formulated. The latter description uses a pore-scale, mechanistic approach for describing rate-limited water and solute distributions, whereas the macroscopic description of solute mass transfer use a lumped–parameter, empirical approach.

The macroscopic approach mathematically describes the interaction of solutes between mobile and immobile pore domains using an empirical first-order expression, where knowledge of the system geometry or pore structure is not necessary. Because of this, the physical significance of the rate coefficient is lost; however, the time-dependency of the solute mass-transfer process is known. The governing transport equations that describe this scenario are expressed as:

$$\theta_m R_m \frac{\partial C_m}{\partial t} + \theta_{im} R_{im} \frac{\partial C_{im}}{\partial t} = \theta_m D_m \frac{\partial^2 C_m}{\partial x^2} - \theta_m v_m \frac{\partial C_m}{\partial x} \qquad [1]$$

$$\theta_{im} R_{im} \frac{\partial C_{im}}{\partial t} = \alpha (C_m - C_{im}) \qquad [2]$$

where subscripts 'm' and 'im' denote mobile and immobile domains, respectively; C denotes the solution concentration of the solute; D is the hydrodynamic dispersion coefficient; v is the average pore-water velocity; $R = R_m + R_{im}$ is the retardation factor; α is the first-order mass-transfer coefficient; $\theta = \theta_m + \theta_{im}$ is the volumetric water content; x is soil distance; and t is time. As written, eqn [1] assumes a linear isotherm of the form $s = k_d C$, where s is the adsorbed solute concentration and k_d is the equilibrium distribution coefficient. Therefore, $R = 1 + (\rho k/\theta)$ where ρ is the porous media bulk density. The partitioning of solutes into mobile and immobile domains is frequently distinguished with a parameter denoted as $F = s_m/s$ that represents the mass fraction of adsorbed solutes that are in direct contact with the mobile liquid phase. The transport model described by eqns [1] and [2] is commonly referred to as the two-region or mobile–immobile model.

The microscopic approach mathematically describes the interaction of solutes between mobile and immobile pore domains using a more rigorous, mechanistic approach that considers the pore geometry and implicit structure of the soil media. Using this approach, the interchange of solute between mobile and immobile zones is often specified by Fick's second law of diffusion, where mass exchange into spherical aggregates and plate-like matrix blocks has been investigated. For uniformly sized spherical aggregates of radius a, eqn [1] remains valid for mobile-phase transport, while eqn [2] is replaced by the spherical diffusion equation:

$$R_{\mathrm{im}} \frac{\partial C_{\mathrm{a}}}{\partial t} = \frac{D_{\mathrm{a}}}{r^2} \frac{\partial}{\partial r}\left(r^2 \frac{\partial C_{\mathrm{a}}}{\partial r} \right) \quad (0 < r < a) \qquad [3]$$

where C_{a} is the solute concentration in the aggregate; r is the radial coordinate; and D_{a} is the effective matrix diffusion coefficient. The mean immobile solute concentration C_{im} in eqns [1] and [2] is now the mean solute concentration of the intraaggregate liquid phase:

$$C_{\mathrm{im}}(x,t) = \frac{3}{a^2} \int_0^a r^2 C_{\mathrm{a}}(x,r,t)\,\mathrm{d}r \qquad [4]$$

Another class of two-region flow and transport models also exist that assume both pore regimes are mobile (i.e. no immobile zones) with one pore-class dominated by advective–dispersive flow and the other pore-class serving primarily as a source/sink for slowly migrating solutes. Interaction between the pore regimes is described by an empirical first-order mass transfer coefficient.

Multiregion approach Since soils are typically a vast continuum of pore regimes of varying size and shape, the two-region approach may oversimplify fate and transport processes in highly structured soils known to possess multiregion flow regimes. Experimental confirmation of multiregion flow and transport in soils has resulted in the development of numerically advanced multiporosity, multipermeability, flow and transport mathematical models. A three-region visual conceptualization of this approach is provided in Figure 1, where a representative elementary volume (REV) at any point in the soil consists of three regions (e.g., macro-, meso-, micropores), each with its own flow and transport parameters. Intraregion mass transfer is described by flow from a physical point to a neighboring one, and interregion mass transfer between the various pore regimes is controlled by both advective and diffusive processes. Advective mass transfer is driven by hydraulic gradients that are caused by differences in fluid velocity in different-sized pores, and diffusive processes are caused by concentration gradients between pore classes. Another process inherent in multiregion flow and transport in structured systems is the time-dependent nature of both the advective and diffusive mass transfer rates between the various pore domains. Time-dependent mass transfer rates take into account changes in concentration gradients as solute mass is transferred between pore regimes. Time-dependent rate coefficients also indirectly account for variabilities in the size of soil aggregates and matrix blocks.

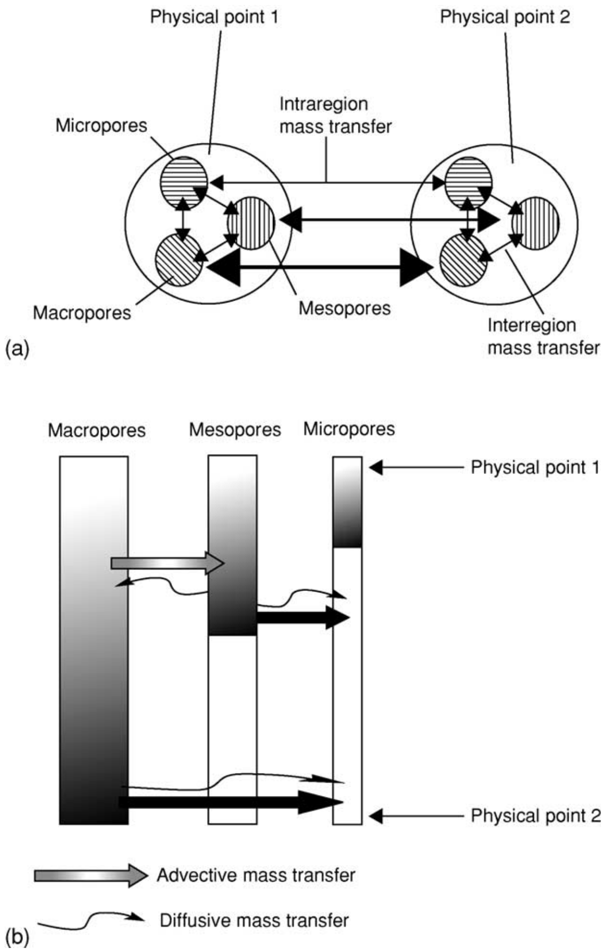


Figure 1 Triple-porosity, triple-permeability example of the multiregion flow and transport concept: (a) the representative elementary volume (REV) (large circle) at two physical points consists of three pore regions. Intraregion flow and transport is indicated by lines between large circles (REVs) and interregion transfer is depicted by lines within each large circle (REV); (b) both advective and diffusive mass transfer may occur in parallel or counter to each other. (Adapted from Gwo JP, Jardine PM, Wilson GV, and Yeh GT (1996) *Water Resources Research* 32: 561–570.)

Chemical and Microbiological Nonequilibrium

The occurrence of chemical and microbiological nonequilibrium during solute transport in soils often depends on the mineral composition and physical makeup of the soil as well as the chemistry and interconnectedness of the pore water. Rate-limited geochemical and microbiological reactions such as adsorption, transformation, precipitation–dissolution, and oxidation–reduction are commonplace in soil systems where transient storm infiltration events inevitably disrupt the equilibrium state of most any interfacial process.

Adsorption Nonequilibrium conditions during solute transport may be caused by kinetically limited adsorption reactions. Frequently, biphasic ion adsorption on soils is encountered where rapid reactions on readily accessible sites occur simultaneously with slow reactions on less-accessible sites. The energetics of adsorption may differ on these two types of sites depending on their location within the structure of the soil mineralogy. Certain locations exhibit structural constraints and increased charge densities (e.g., interlayer spaces) that influence the rate of chemical reaction during solute transport. Often, the more restrictive interlayer spaces of clay minerals exhibit larger solute adsorption energetics relative to easily accessible external mineral surfaces.

Disequilibrium due to chemical kinetic limitations on heterogeneous soil surfaces have been modeled using a two-site nonequilibrium approach, which assume solute adsorption on the two types of sites occur at different rates. Generally, empirical first-order and second-order expressions are utilized to describe the nonequilibrium adsorption process. The governing transport equation that describes solute adsorption on two different types of sites may be expressed as:

$$\frac{\partial C}{\partial t}+\frac{\rho}{\theta}\left(\frac{\partial s_1}{\partial t}+\frac{\partial s_2}{\partial t}\right)=D\frac{\partial^2 C}{\partial x^2}-v\frac{\partial C}{\partial x} \qquad [5]$$

where parameter definitions are analogous to those in eqn [1] and s_1 and s_2 are the adsorbed solute concentrations for type-1 and type-2 sites, respectively. The rate of solute interaction with the two different sites can be described by a variety of kinetic expressions listed below:

$$\text{Lumped first-order} \quad \frac{\partial s_{\mathrm{i}}}{\partial t}=\alpha_{\mathrm{i}}\left(s_{\mathrm{i}}^{*}-s_{\mathrm{i}}\right) \qquad [6]$$

$$\text{Reversible first-order} \quad \frac{\partial s_{\mathrm{i}}}{\partial t}=\alpha_{\mathrm{bi}}\,s_{\mathrm{i}}^{*}-\alpha_{\mathrm{fi}}\,s_{\mathrm{i}} \qquad [7]$$

$$\text{Second-order} \quad \frac{\partial s_{\mathrm{i}}}{\partial t}=\alpha_{\mathrm{fi}}\psi_{\mathrm{i}}\,C-\alpha_{\mathrm{bi}}\,s_{\mathrm{i}} \qquad [8]$$

$$\text{General-order} \quad \frac{\partial s_{\mathrm{i}}}{\partial t}=\alpha_{\mathrm{fi}}\,C^{n}-\alpha_{\mathrm{bi}}\,s_{\mathrm{i}} \qquad [9]$$

where s_{i}^{*} and s_{i} are the concentrations of adsorbed solute at equilibrium and any time t for type i sites, respectively; α_{f} and α_{b} are forward (adsorption) and backward (desorption) first-order rate coefficients for type-i sites, with $i=1$ and 2 for type-1 and type-2 sites, respectively; ψ_{i} is the amount of vacant sites of type i; and n is the moles of reactant solute in solution, the 'reaction order.' Thus, both adsorption sites can be kinetically controlled or type-1 sites can be at equilibrium ($\alpha_1 \rightarrow \infty$) while type-2 sites remain kinetically controlled.

Transformation Nonequilibrium behavior during solute transport in soil may also result from time-dependent chemical and biological transformation reactions. Consideration of chemical fixation, dissolution, hydrolysis, and polymerization reactions in mathematical models are often necessary to correctly describe the transport of certain solutes in soil. Time-dependent biological reactions, which transform solutes into a variety of chemical species, may also need consideration during solute transport in some soils. Nitrification and denitrification processes that result from the application of N fertilizer sources to soils, are examples of complex subsurface processes that are influenced by time-dependent chemical and microbial reactions. Nitrogen fertilizers such as urea are rapidly hydrolyzed by the enzyme urease to form NH_4. The ammonium ion is a reactive cation in soils, which can become fixed within interlamellar spaces of certain 2:1 phyllosilicates, and is subject to microbial oxidation, resulting in the formation of NO_3. This oxidation reaction is known as nitrification and is typically governed by either zero-order or first-order kinetics. The microbial-mediated reduction of NO_3 (denitrification) also occurs in soils, resulting in time-dependent formation of N_2 and N_2O gases. Numerous transport models have been adapted to consider nonequilibrium nitrogen transformation reactions in soil during one- and three-dimensional flux of soil water.

Soil reactions involving potassium fixation and release can also be considered a time-dependent transformation process, since soil K^+ can be partitioned between solution, exchangeable, nonexchangeable (fixed), and primary mineral phases (Figure 2). Transformation of K from one phase to another is induced by transient soil conditions that promote K mineral weathering, leaching losses of soil solution K, and plant uptake of soil solution and exchangeable K. Exchangeable K is largely bound on readily accessible soil sites which may or may not be in local equilibrium with the solution phase. Many studies have shown that the adsorption and release of exchangeable K may be kinetically limited in the presence of 2:1 nonexpanding clay minerals. Nonexchangeable and mineral K retention and release reactions in soil are typically very time-dependent

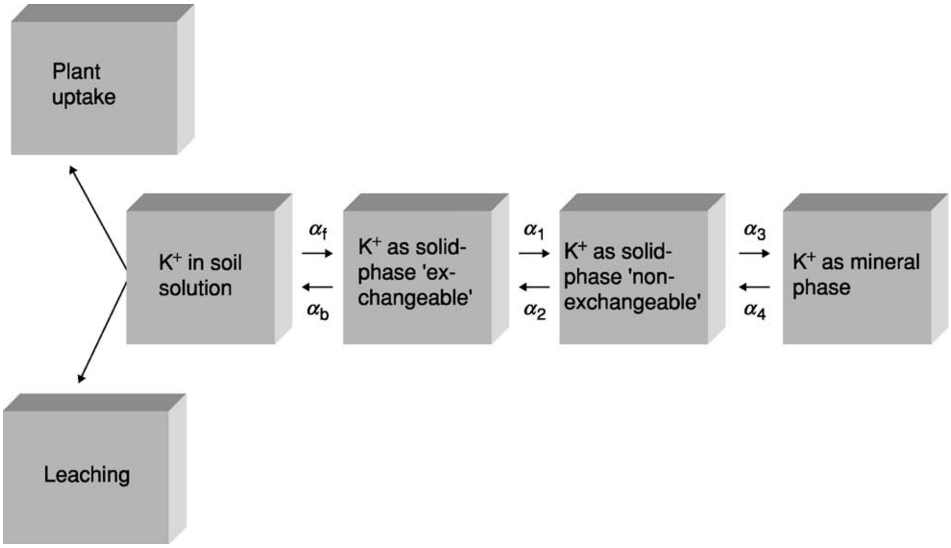


Figure 2 Schematic representation of the time-dependent reactions of potassium in soil. (Adapted from Selim HM, Mansell RS, and Zelazny LW (1976) Modeling reactions and transport of potassium in soils. *Soil Science* 122: 77–84.)

relative to transformations between solution and exchangeable K phases. Modeling K transport in soils, therefore, requires consideration of the nonequilibrium transformation reactions which occur between the various K phases.

A mathematical formulation of the conceptual model described in Figure 2 regarding the kinetics of K transformation during transport is described below. Transformations in the exchangeable phase (denoted by subscript 'ex') are represented by:

$$\frac{\partial s_{\text{ex}}}{\partial t} = \left(\frac{\theta}{\rho}\right)\alpha_f C^n + \alpha_2 s_{\text{nex}} - (\alpha_1 + \alpha_b)s_{\text{ex}} \qquad [10]$$

where parameter definitions are synonymous with eqns [1] and [5], α_f and α_b are the forward (adsorption) and backward (desorption) rate coefficients for reactions between the solution and exchangeable phases, α_1 and α_2 are the nonexchangeable fixation and release rate coefficients, respectively, and n is the order of the reaction for adsorption of K from solution to the exchangeable phase. Transformations in the nonexchangeable phase (denoted by subscript 'nex') is defined by:

$$\frac{\partial s_{\text{nex}}}{\partial t} = \alpha_1 s_{\text{ex}} + \alpha_4 s_{\text{min}} - (\alpha_2 + \alpha_3)s_{\text{nex}} \qquad [11]$$

where α_3 and α_4 are the immobilization and the mineralization rate coefficients describing time-dependent K reactions between nonexchangeable and mineral phases, and s_{min} is the amount of K in the mineral phase. Finally, transformations in the mineral phase are represented by:

$$\frac{\partial s_{\text{min}}}{\partial t} = \alpha_3 s_{\text{nex}} - \alpha_4 s_{\text{min}} \qquad [12]$$

Based on eqns [10–12], it can be assumed that the reactions between exchangeable and nonexchangeable as well as those between nonexchangeable and primary minerals are first-order kinetic reactions. Coupling eqns [10–12] with the one-dimensional convective–dispersive equation (one-site version of eqn [5]), one can describe the transformation kinetics of K during transport in soil.

The polymerization of solutes in soil, such as Al, may also be considered a time-dependent transformation process. The significance of Al transport through subsurface soil environments is of great concern to agriculturalist and environmentalist. Soil-solution Al is known to adversely effect terrestrial and aquatic ecosystems by severely limiting plant and fish production. It has also been linked to several human health disorders such as Alzheimer's disease and senile dementia. Therefore, the biogeochemical and hydrologic processes controlling the transport and cycling of Al are of considerable economic and scientific interest. The polymerization of Al tends to alleviate the adverse toxicity effects associated with Al^{3+}. Soil surfaces have been shown to serve as template for enhanced Al polymerization and these reactions have been modeled as multisite, first-order limited adsorption processes.

Solute transport in soils may be complicated by precipitation reactions which act as kinetic sinks for chemicals. Because of its agronomic and environmental significance, phosphorus (P) precipitation reactions during transport in soil have been most frequently studied. Often P is a severely limited nutrient in terrestrial and aquatic ecosystems which adversely affects plant and fish production. Because of the significance of P precipitation reactions in soil, the biogeochemical and hydrologic processes controlling P transport must be clearly understood. Many of the principles associated with P reactions may in some instances be applicable to other types of precipitation reactions (i.e., Al, Fe, etc). Soil reactions involving P are generally modeled as first-order irreversible precipitation reactions described by:

$$Q = \alpha_c(\theta C) \qquad [13]$$

and first-order irreversible chemical immobilization reactions described by:

$$Q = \alpha_s(\rho s) \qquad [14]$$

where α_c and α_s are the rate coefficients for precipitation and chemical immobilization, respectively, and Q is the rate of solute consumption (sink). These kinetic sinks for P are often coupled to convective–dispersive transport with consideration of time-dependent, nonlinear adsorption to the soil surface. Certain modeling endeavors also consider the rate of P nucleation and crystal growth based on solubility product theory.

Biotransformation reactions in soils are often kinetically limited due to time-dependent microbial growth rates. Biologically mediated detoxification, mineralization, and electron-transfer reactions involving organic or inorganic solutes are often time-dependent processes that are mathematically described by either empirical zero- and first-order reactions, or Monod-type kinetics that account for microbial growth rates. Simple Monod kinetics consider the loss of substrate concentration, s (e.g., organic or inorganic pollutants) as a function of microbial growth rate, μ, where:

$$\mu = \frac{\mu_{\max}[s]}{k_s + [s]} \qquad [15]$$

with $\mu_{\max}$ is maximum specific growth rate, and k_s is substrate 0.5 saturation constant. When multiple substrates are available as an energy source for microbial growth, double Monod kinetics have been utilized, where a second substrate, the electron acceptor O_2, is used as here:

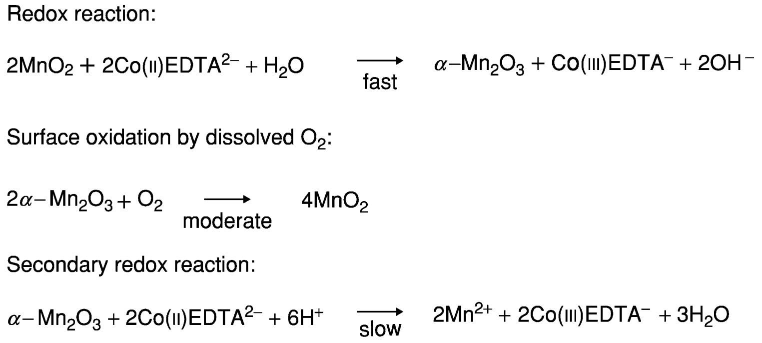
$$\mu = \left[\frac{\mu_{\max}[s]}{k_s + [s]}\right]\left[\frac{\mu_{\max,O_2}[O_2]}{k_{O_2} + [O_2]}\right] \qquad [16]$$

Monod kinetics have also been incorporated into rate expressions for substrate loss:

$$\frac{-d[s]}{dt} = \mu\frac{X}{Y} = \mu_{\max}\frac{[s]}{k_s + [s]}\frac{X}{Y} = k_b\frac{[s]X}{k_s + [s]} \qquad [17]$$

where Y is yield (mass/mass), X is biomass per volume, and k_b is the biodegradation rate constant (t^{-1}). Any one of these expressions have been coupled with flow and transport equations in an effort to model time-dependent biotransformation processes in flowing, dynamic systems indicative of soil profiles.

Oxidation–Reduction Redox reactions at the solid–solution interface typically alter the geochemical behavior of many environmentally significant metals and organics. Soil solid-phase constituents often accelerate redox related charge transfer reactions and thus catalyze the oxidation or reduction of numerous redox-sensitive solutes. For example, common soil constituents such as Mn(III/IV) oxides and Fe(III) oxides are powerful oxidants that catalyze the oxidation of metals such as Co(II), Cr(III), and As(III). Other soil constituents such as organic matter and Fe(II)-bearing minerals, are known to be strong reductants with regard to metals such as Co(III) and Cr(VI). Models have been developed that couple transport and oxidation–reduction, and these models have been tested on systems involving redox-sensitive contaminants such as Cr and Co in the presence of mineral oxidants and reductants. As an example, the complex oxidation reactions of Co(II)-ethylenediaminetetraacetic acid (EDTA) by Mn oxides during transport have been modeled as shown in Figures 3 and 4. In this study, macroscopic transport experiments, coupled with the real-time spectroscopic interfacial interrogation methods, reveal that Mn(IV) oxides catalyze the time-dependent oxidation of Co(II)EDTA^{2-} to Co(III)EDTA^{-} (Figures 3 and 4). The redox reaction produces an intermediate Mn(III)-oxide phase, α-Mn_2O_3, on the surface of the Mn(IV) mineral that significantly passivated the redox reaction. Time-dependent reoxidation of Mn(III) to Mn(IV) is controlled by dissolved oxygen (DO). Successful modeling of these reactions involves coupling the advective–dispersive transport of Co(II)EDTA^{2-}, Co(III)EDTA^{-}, and DO with multiple, nonlinear kinetic expressions that describe: (1) oxidation of Co(II)EDTA to Co(III)EDTA by β-MnO_2, (2) inhibition of Co(II)EDTA oxidation due to precipitation of Mn_2O_3, an insoluble reaction product that occludes the β-MnO_2 surface, and (3) regeneration of the redox-reactive β-MnO_2 surface through oxidation of the Mn_2O_3 precipitate by dissolved oxygen (Figure 4). Comparison of experimental and calculated results demonstrates that the model describes the response of the coupled hydrologic and geochemical process to change in



Redox reaction:

$$2MnO_2 + 2Co(II)EDTA^{2-} + H_2O \xrightarrow[\text{fast}]{} \alpha\text{-}Mn_2O_3 + Co(III)EDTA^{-} + 2OH^{-}$$

Surface oxidation by dissolved O_2:

$$2\alpha\text{-}Mn_2O_3 + O_2 \xrightarrow[\text{moderate}]{} 4MnO_2$$

Secondary redox reaction:

$$\alpha\text{-}Mn_2O_3 + 2Co(II)EDTA^{2-} + 6H^{+} \xrightarrow[\text{slow}]{} 2Mn^{2+} + 2Co(III)EDTA^{-} + 3H_2O$$

Figure 3 Spectroscopically confirmed reaction sequence describing the time-dependent, multireaction oxidation of Co(II)EDTA^{2-} to Co(III)EDTA^{-} in the presence of Mn oxides. (Adapted from Fendorf SE, Jardine RR, Patterson RR, Taylor DL, and Brooks SC (1999) Pyrolusite surface transformations measured in real-time during the reactive transport of Co(II)EDTA. *Geochimica et Cosmochimica Acta* 63: 3049–3057.)

flow rate, influent concentrations of Co(II)EDTA, and β-MnO_2 surface coverages.

Multicomponent, Multiprocess Models

In a system of multiple components, the identity of a component becomes chemically distinct from that of a species. Components are a set of linearly independent 'basic' chemical entities such that every species can be uniquely represented as a combination of those components. A species is then defined as the product of a chemical reaction, involving the components as reactants. Thus, numerical models that consider multiple components and multiple processes are computationally intensive, especially when aqueous- and solid-phase reaction kinetics are considered. The complexity of such models is often driven by the complexities observed in nature where enhanced interactions between experimentalist, mathematicians, modelers, and decision makers ultimately dictate the necessary processes that are needed to represent the desired system.

Nonequilibrium Processes Involving Dissolved Aqueous Phase Constituents

In recent years, a limited number of multicomponent reactive transport models have become available that couple hydrodynamic transport with multiprocess, time-dependent geochemical and microbial reactions. The equilibrium-based, multicomponent model HYDROGEOCHEM has been modified (HYDROGEOCHEM 2.0) to simulate mixed chemical kinetic and equilibrium reactions in an effort to deal with multispecies fate and transport that is controlled by time-dependent or thermodynamic-based reactions. Further modifications have incorporated time-dependent microbial reactions (HYDROBIOGEOCHEM), designed to solve a system of equations describing hydrologic transport and biogeochemical reactions in a reactive, multicomponent unsaturated or saturated system. The major transport processes are advection, dispersion–diffusion, and source–sinks. The major chemical processes are aqueous complexation, adsorption, ion-exchange, precipitation–dissolution, redox, and acid–base reactions. The major microbiological processes are biodegradation and microbial respiration. Rate expressions for kinetically controlled geochemical reactions are based on collision theory and can be represented generally by a stoichiometric equation of the following form:

$$\sum_{j\in M_x} v_j c_j \leftrightarrow \sum_{j\in M_x} \mu_j s_j \qquad [18]$$

where c_j and s_j are the reactants and products, respectively; v_j and μ_j are the reactant and product stoichiometric coefficients, respectively; and M_x is the set of chemical species of interest. A generic rate equation corresponding to eqn [18] may then be written as:

$$\Omega = \alpha_f \prod_{j\in M_x} \left(\gamma_j c_j\right)^{vj} - \alpha_b \prod_{j\in M_x} \left(\gamma_j s_j\right)^{\mu j} \qquad [19]$$

where Ω, α_f, and α_b are the reaction rate, forward rate constant, and backward rate constant, respectively, and γ_j is the activity coefficient of species j according to the Davies formula. Microbial growth rates are represented by a modified Monod kinetic equation as follows:

$$\Omega = \Gamma I_g \left[\frac{S}{(k_s I_s + S)}\right] \left[\frac{A}{k_a + A}\right] \left[\frac{N}{k_n + N}\right] L \qquad [20]$$

where Γ is the microbial growth rate constant; k_s, k_a, and k_n are the half-saturation constants of substrate, electron acceptor, and nutrient, respectively; I_g and I_s are the inhibition factors for the growth rate and the half-saturation constant of the substrate, respectively; L is the metabolic lag coefficient; and S, A, and N are

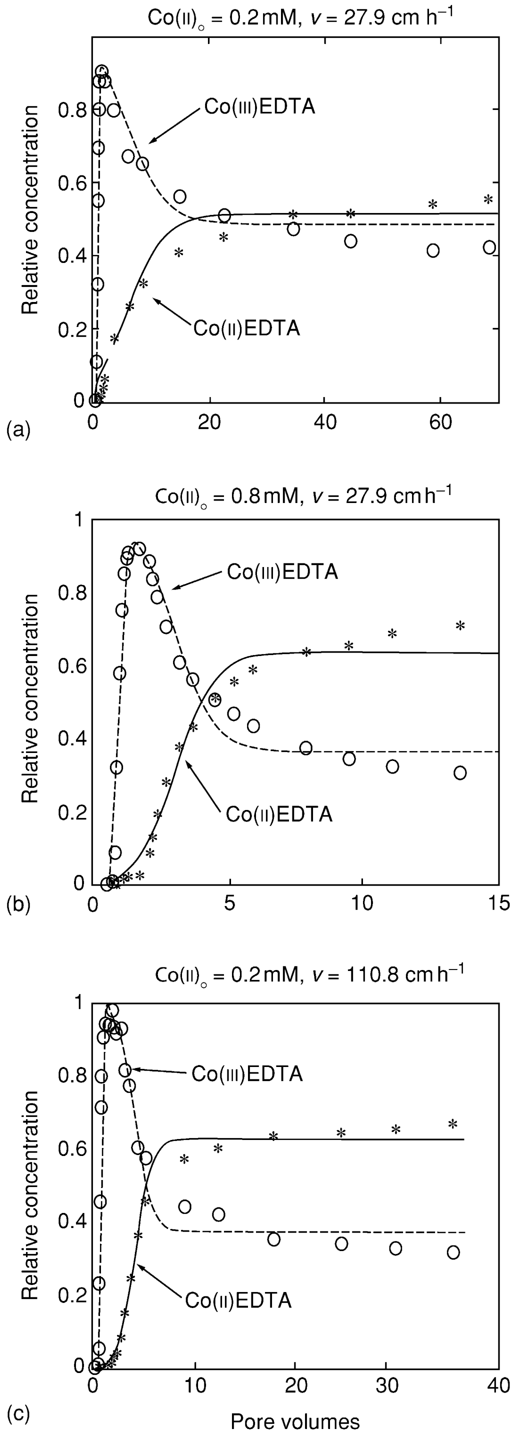


Figure 4 Observed and modeled fitted breakthrough curves for the displacement and subsequent oxidation of Co(II)EDTA^{2-} in columns of Mn(IV) oxide-coated sand for different influent concentrations and residence times. The model considered multiple nonlinear kinetic expressions to account for the complex series of oxidation and reduction reactions. (Reproduced from Saiers JE, Guha H, Jardine PM, and Brooks SC (2000) Development and evaluation of a mathematical model for the transport and oxidation–reduction of CoEDTA. *Water Resources Research* 36: 3151–3165.)

the concentrations of substrate, electron acceptor, and nutrient, respectively.

A high-performance computing version of HYDROBIOGEOCHEM has also been released (HBGC123D) which is a three-dimensional, shared-memory, parallel computer model that couples hydrogeologic and biogeochemical processes. The parallelized code offers the advantage of simulating large-scale, more realistic problems with improved performance and speed (Figure 5). Other multicomponent reactive transport models such as CRUNCH and FERACT are also available which couple multiple sets of kinetically controlled geochemical and biological processes with advective–dispersive flow. Still other kinetic-based flow models are available that target specific problems related to time-dependent, microbially driven redox reactions and complex, time-dependent metalorgano geochemical reactions in heterogeneous subsurface environments.

Nonequilibrium Processes Involving Dissolved Nonaqueous-Phase Constituents

Numerous multicomponent reactive transport models involving nonaqueous-phase constituents have also been developed that couple hydrodynamic transport with multiprocess, time-dependent geochemical and microbial reactions. The multiphase flow and multicomponent reactive transport simulator PARSim has been linked to a mixed chemical kinetic and equilibrium model (KEMOD) to allow simulation of multiple flowing phases with a full complement of reactions. The model (RPARSim/KEMOD) allows for a more general, nonequilibrium phase-transfer for KEMOD-style reactions where the reactants and products are in different phases. The model has been parallelized in order to enhance computational efficiency and the need to simulate larger-scale, more-realistic environmental problems. Other kinetic-based models designed to deal with subsurface dense, nonaqueous-phase liquid (DNAPL) issues, tend to emphasize substrate-limited biodegradation. The US Environmental Protection Agency (EPA) code BIOPLUME III is a two-dimensional contaminant transport model that couples DNAPL advective–dispersive transport with sorption, first-order decay, and biodegradation through instantaneous, zero-order, first-order, or Monod kinetics. The model is based on the US Geological Survey (USGS) BIOMOC code, where the hydrocarbon source and each active electron acceptor (O_2, NO_3^-, Fe(III), SO_4^{2-}, and CH_4) are simulated as separate plumes.

Another code, 'MT3D99,' which is an enhancement of MODFLOW, also couples advective–dispersive transport in soil systems with nonequilibrium sorption, time-dependent, nonaqueous-phase liquid dissolution,

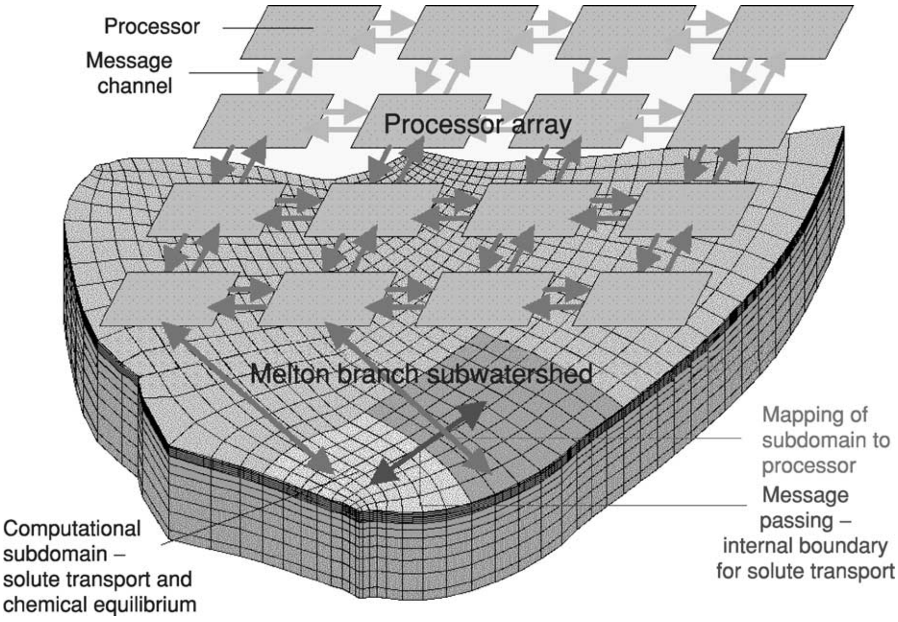


Figure 5 How solute fate and transport processes are modeled at the watershed scale using experimental subsurface transport facilities at the Oak Ridge National Laboratory. Computational subdomains are mapped to processors on a massively parallel computer. Since watershed-scale models are built with imperfect information of the subsurface and uncertainties associated with model parameters and processes, high-performance computing is necessary for quantifying parameter uncertainties when using mechanistically rigorous transport models at the field scale. Parallel-processing supercomputers provide the capability of solving large, realistic problems fast and economically. (Adapted from Gwo JP, Wilson GV, Jardine PM, and D'Azevedo EF (1999) Modeling subsurface contaminant reactions and transport at the watershed scale. Proceedings of the 1997 Chapman/Outreach Conference 'Assessment of Non-Point Source Pollution in the Vadose Zone.' Geophysical Monograph 108, American Geophysical Union, pp. 31–43.)

and rate-limited microbial processes. It considers Bioplume-type reactions, monad reactions, and daughter products, thus enabling the simulation of multispecies transport in a similar manner to the transport code RT3D. Kinetically limited hydrocarbon biodegradation using multiple electron acceptors and time-dependent transport of bacteria, electron acceptors, and hydrocarbons are explicitly considered. MT3D99 also maintains a dual-porosity option, where the soil media is divided into an advective-dominated mobile domain and a diffusion-dominated immobile domain. An empirical, first-order parameter accounts for mass transfer between the domains.

Although some multicomponent, multiporosity models do exist, typically they are limited by a small number of finite components. Full-blown, computationally rigorous, multicomponent, multiporosity, multipermeability models that consider coupled, time-dependent physical, chemical, and microbial reactions, have not yet been formulated or numerically implemented as of 2003. The utility of such models for simulating kinetically limited subsurface processes is conceptually appealing; however, parameter uncertainty resulting from limited experimental data at appropriate scales remains an obstacle for their practical use.

See also: **Cation Exchange**; **Chemical Equilibria**; **Surface Complexation Modeling**

Further Reading

Arbogast T, Bryant S, Dawson C *et al.* (1996) Computational methods for multiphase flow and transport problems arising in subsurface contaminant remediation. *Journal of Computational and Applied Mathematics* 74: 19–32.

Gwo JP, Jardine PM, Wilson GV, and Yeh GT (1995) A multiple-pore-region concept to modeling mass transfer in subsurface media. *Journal of Hydrology* 164: 217–237.

Gwo JP, D'Azevedo EF, Frenzel H *et al.* (2001) HBGC123D: A high performance computer model of coupled hydrological and biogeochemical processes. *Computers and Geosciences* 27: 1231–1242.

Hunter KS, Wang Y, and Van Cappellen P (1998) Kinetic modeling of microbially driven redox chemistry of subsurface environments: coupling transport, microbial metabolism and geochemistry. *Journal of Hydrology* 209: 53–80.

Hutson JL and Wagenet RJ (1995) A multiregion model describing water flow and solute transport in heterogeneous soils. *Soil Science Society of America Journal* 59: 743–751.

Jardine PM, Parker JC, and Zelazny LW (1985) Kinetics and mechanisms of aluminum adsorption on kaolinite using a two-site nonequilibrium transport model. *Soil Science Society of America Journal* 49: 867–873.

Saiers JE, Guha H, Jardine PM, and Brooks SC (2000) Development and evaluation of a mathematical model for the transport and oxidation-reduction of CoEDTA. *Water Resources Research* 36: 3151–3165.

Selim HM, Mansell RS, and Zelazny LW (1976) Modeling reactions and transport of potassium in soils. *Soil Science* 122: 77–84.

Steefel CI and Lichtner PC (1998) Multicomponent reactive transport in discrete fractures. I. Controls on reaction front geometry. *Journal of Hydrology* 209: 186–199.

Sudicky EA and Frind EO (1982) Contaminant transport in fractured porous media: analytical solutions for a system of parallel fractures. *Water Resources Research* 18: 1634–1642.

Szecsody JM, Zachara JM, Chilakapati A, Jardine PM, and Ferrency AS (1998) Importance of flow and particle-scale heterogeneity on Co(II/III)EDTA reactive transport. *Journal of Hydrology* 209: 112–136.

Tebes-Stevens C, Valocchi AJ, Van Briesen JM, and Rittmann BE (1998) Multicomponent transport with coupled geochemical and microbiological reactions: model description and example simulations. *Journal of Hydrology* 209: 8–26.

Wagenet RJ, Biggar JW, and Nielsen DR (1977) Tracing the transformations of urea fertilizer during leaching. *Soil Science Society of America Journal* 41: 896–902.

Yeh GT and Tripathi VS (1991) A model for simulating transport of reactive multispecies components – model development and demonstration. *Water Resources Research* 27: 3075–3094.

Yeh GT, Salvage KM, Gwo JP, Zachara JM, and Szecsody JE (1998) HYDROBIOGEOCHEM: a coupled model of HYDROlogic transport and mixed BIOGEOCHEMical kinetic/equilibrium reactions in saturated-unsaturated media. ORNL/TM-13668. Oak Ridge TN: Oak Ridge National Laboratory.

KIRKHAM, DON

D R Nielsen, University of California, Davis, CA, USA
R R van der Ploeg, University of Hannover, Germany

Don Kirkham, American soil physicist and educator who made pioneering and fundamental contributions in solving soil water flow problems relevant to agriculture, was born in Provo, Utah, on February 11, 1908. He was the third child of six living children of Francis Washington and Alzina Robison Kirkham. His father, an educator from a Mormon family, had been an early missionary in New Zealand and wrote the first Maori grammar. The family moved to Salt Lake City in 1911, where the father served as president of a business college. In Salt Lake City, Don Kirkham attended Emerson School, and although absent a great deal of the time due to various children's diseases, he did well. His teachers recognized his ability and he skipped sixth grade. In the summers of 1919, 1920, 1921, 1922, and 1925, Kirkham worked at the family's West Jordan farm, incidentally acquiring experience to be useful in his later professional career in soil physics.

In December 1922, his family moved to Berkeley, California, for a stay of 1.5 years while his father worked for his PhD degree. Don attended Edison Junior High School (where he was elected vice-president of his class) and Berkeley High School. He graduated from East High School in Salt Lake City in June 1925, completing the scientific course. In the autumn of 1925, while attending the University of Utah, he developed Perthes disease (bone disease). His hip and left leg were kept in a cast from January 1926 to August 1926. He had to drop out of the University of Utah because it was not equipped for people on crutches. However, Kirkham had always had a deep interest in music. For Christmas when he was only 11 years old, he was given a clarinet and started music lessons with Lt Clarence J. Hawkins, just returned from World War I as director of the 145th Field Artillery band. Kirkham graduated from

Figure 1 Don Kirkham *c.* 1985.

the McCune School of Music and Art in Salt Lake City. When he graduated from the McCune School as teacher and soloist of clarinet, he played his graduation recital while in the cast. In the theory of music at this institution, he was high man in the final examination. In the summer of 1926, he joined the Musicians Union, the American Federation of Musicians, and earned considerable money playing and arranging band music and even composed a band number, the 'Ute Patrol.'

In 1927, he was called on a mission to Germany by Heber J. Grant, president of the Mormon Church. In Germany, he served as president of the Hamburg district, the largest district in point of membership in the Swiss-German mission. He had the reputation of knowing German grammar better than any others in the mission. At the end of his mission in 1930, his mother sent him the sum of $600, which had been left to her by an uncle. Kirkham used the money to travel extensively and frugally throughout Europe.

After his travels, he arrived in New York City in the autumn of 1930, where his family was then living, his father being the educational director of the National Child Welfare Association. He entered Columbia College that autumn. Three years later he graduated with honors in Physics. In the meantime he had won a Columbia College Scholarship and held scholarships in the Columbia Concert Band. In 1934, he received the Master's degree and in 1938 the PhD degree. While working on the doctor's degree, he was Assistant in the Ernest Kempton Adams (Physics) Precision Laboratory at Columbia University.

In September 1937, he arrived at Logan, Utah, where he had accepted a position to teach in the Physics and Mathematics Departments of the State Agricultural College. Three years later, as an Assistant Professor, he resigned to accept a position with the Navy Department, Bureau of Ordnance. The Navy was in need of physicists to protect ships against magnetic mines then employed by the Germans against British ships. He aided, first at Yorktown, Virginia, and then in New York City Harbor, in the development of a process for removing the permanent magnetism from ships called deperming, and in other techniques of degaussing, which was the general name employed for processes reducing ships' magnetism. He was in charge of the design, construction, and operation of degaussing facilities at New York Harbor. In January 1943 he was transferred to the Naval Ordnance Laboratory at Washington, DC, where he worked in torpedo research until September 1946.

From May 1946 to the middle of August of the same year, he was a civilian member of the scientific staff of Operations Crossroads, the Joint Army–Navy Task Force One. The mission of this Task Force was to test the atom bomb at Bikini Atoll in the Marshall Islands. Two bombs were exploded, one an air blast, the other an underwater blast. He witnessed both blasts from the USS Kenneth Whiting, a few miles outside the atoll. He went to Bikini by boat, being Senior Physicist of a group of Naval Ordnance Laboratory men, and was in the boarding party of scientists after the second and most powerful shot.

After his return to the USA from Bikini, he joined the staff at Iowa State College (now Iowa State University), one of the nation's leading agricultural schools, as Research Associate Professor in Soil Physics. His interest in soil physics had originated at Utah State Agricultural College, where he had been encouraged by Dr Willard Gardner, the eminent soil physicist there.

In 1938, Kirkham met and later married Mary Elizabeth ('Betty') Erwin Kirkham, who had taught English at Utah State. They had three children: Victoria, a professor in the Department of Romance Languages at the University of Pennsylvania; Mary Beth, a professor in the Department of Agronomy at Kansas State University; and Don Collier, a corporate lawyer and partner in the law firm of Cravath, Swaine & Moore in New York City.

Kirkham won many honors. He held the title of Charles F. Curtiss Professor of Agriculture at Iowa State University. He was the only professor to hold joint professorships in both the Department of Physics and the Department of Agronomy. He was a Fellow of the American Society of Agronomy, the American Physical Society, the Soil Science Society of America, and the American Geophysical Union. He was named Honorary Member of the Soil Science Society of America, the International Soil Science Society, the International Soil Tillage Research Organization, and the American Water Resources Association. He was a Fulbright-Hays Program grantee in the Netherlands, 1950–1951, and in Belgium, 1957–1958.

He received a Guggenheim Award to study in Belgium, 1957–1958. He was awarded an Honorary Doctor of Agriculture Degree from the Royal Agricultural University, Ghent, Belgium, in 1963 and an Honorary Doctor's Degree from Ohio State University in 1993. He was the First Inductee of the Drainage Hall of Fame at Ohio State University in 1979. He received the Horton Medal of the American Geophysical Union in 1995 and was a co-winner with Cornelius de Wit of the Netherlands of the 1983–1984 Wolf Foundation Prize for Agriculture. The citation read for 'innovative contributions to the quantitative understanding of soil, water and other environmental interactions influencing crop growth and yield.'

Rendering service to his profession, he traveled extensively and spent extended periods of time in the Netherlands, Belgium, Egypt, and Turkey. He served on the editorial boards of both *Soil Science* and *Water Resources Research*. In addition to his exceptional abilities in research and teaching, Kirkham was a man of outstanding character and integrity. He was gentle, generous, and unpretentious, and recognized, without envy, skills and achievements of others. Don Kirkham died in Ames, Iowa, on March 7, 1998.

Kirkham's Scientific Contributions

Kirkham's early experiments at Utah State on land drainage pertained to land overlying an artesian aquifer under pressure. His results showed that spacing of horizontally laid drain tiles over such an area would not be economical. Vertically pumped drainage wells should be used. He made sand-tank models showing how the artesian drainage occurred and demonstrated that the seepage patterns depended on the boundary conditions rather than on whether the soil was a clay or a sand or a more permeable soil.

He next considered tile drainage of land over an impervious layer. He obtained equations for the streamlines, or water flow paths, and matched them against the sand-tank model flow paths obtained with dyes. A striking discovery of the theory was that by placing a drain line deeper and closer to an impervious layer, the flow would be less (rather than more) when compared with the flow from a less deep drain line in the soil. Details and amplified theory of this problem were published after World War II, in a paper which, although one in agriculture, received the 1948 mathematical award of the Iowa Section of the American Mathematical Society. The next year his paper on the seepage of water into drainage ditches was designated the best paper of the Iowa Academy of Science in 1949. In this paper he derived theoretically the amount of water that would flow into the drainage ditches. He gave the result in an explicit formula; other researchers in the past had failed to find an explicit formula.

In an early paper in 1945, Kirkham had attacked the problem of simultaneous upward and downward seepage flow into tile drains with excellent agreement of theoretical streamlines with experimental streamlines. But this paper, like others up to then, was for homogeneous soil. So Kirkham then analyzed the problem of nonhomogeneous soil and, with a resulting two-part paper, published in 1951, he received the American Geophysical Union's citation for the best paper in hydrology in 1951.

These earlier papers were for ponded water, that is, completely saturated soil, and thus represented an extreme condition. He next considered the more common condition of a curved water table resulting from successive recharges of water as from irrigation, or as from successive rains. By first neglecting the hydraulic friction in the soil under the water-table arch region and then later correcting for this neglect, he derived an explicit relation for the height of water table to be found between drain lines at certain depths and spacings. He compared his theoretical results with field data and found good agreement. The work was published in the *Transactions of the American Geophysical Union* in 1958. In a paper with a graduate student (the late Sadik Joksoz) as senior author, the solution to this drainage problem was put in a graphical form so that the various physical factors that entered into the problem could be considered in the graph and the solution obtained graphically for the depth and spacing of drains. This same student, in working for his PhD degree, considered stratified soil. Two other papers resulted, in which the ratio of hydraulic conductivities of a layered soil could be taken into account. This long, two-part paper was published in the Irrigation and Drainage Division of the American Society of Civil Engineers.

All the foregoing work was for a simplified steady-state condition of either ponded water or steady recharge to the drain tubes. Kirkham next analyzed the problem of a falling water table. He found equations that predicted where the water table arch would be after a certain time when the steady recharge was discontinued. He compared his theoretical results with actual field data that he had obtained with other scientists in the Netherlands many years before (in 1950–1951) and with other field data and published the work in 1964. There was excellent agreement.

In all the above work involving tile drains, it was assumed that the water entered the drain tube uniformly over its length. This would be true if there were an envelope of coarse material around the drain tube. Actually drain pipes were about 30–60 cm long and had about 1–6 mm spacing between adjacent drain pipes. (At least this had been the condition up until recent times when plastic tubes, perforated to let the water enter, became prevalent.) To see just how much the gaps between individual tile lengths or the perforations in tubes would control drainage flow if there were not a pervious envelope around them, Kirkham authored or co-authored several articles. A first one, published in 1951 in two parts, was coauthored by one of his joint soil physics–agricultural engineering PhD candidates (the late Glen O. Schwab), who conceived the idea of using plastic flexible tubing instead of clay pipes in soil. The two-part paper on the influence of perforations on drainage flow to drain tubes received a citation of merit from the American Society of Agricultural Engineers. Perforated flexible plastic tubing is now used almost universally in drainage. Kirkham's theoretical work on the effect of perforations, cracks, and other openings in drain tubes continues to be of great importance.

In an artesian basin in Utah, it had been observed that tile drains would work in certain areas where it might be predicted that such drainage would be uneconomical. Kirkham reasoned that the drainage could be economical, provided there existed an upper layer of soil near the surface that was highly permeable compared with an existing stratum of soil overlying the artesian gravel. With a student and a colleague (Mohammed Najmaii and MD Dougal), he analyzed this problem in a paper published in the Irrigation and Drainage Division of the American Society of Civil Engineers in 1978. Just how much more permeable the upper soil layer must be, compared with that overlying the aquifer, was defined in the paper, which contains numerous graphs of a number of situations to be used in practical application.

All the above papers on water movement into horizontal or vertical drainage pipes depend on a knowledge of a soil coefficient called the hydraulic conductivity. This coefficient gives the ease with which the soil conducts water under an energy gradient. Early, Kirkham recognized that the use of soil cores would not give the hydraulic conductivity accurately because the samples were too small and because of their disturbed condition. He reasoned that, instead of measuring the hydraulic conductivity of soil in a core removed from the soil, measuring the conductivity of the soil around the location where the core was removed or its equivalent would be a better procedure. A practical way to proceed would be to auger a hole into the soil, while simultaneously driving a casing into the hole to seal off the walls of the hole against leakage, except for a cavity at its bottom. Flow would be allowed to come into the cavity, and, by rating this flow, the conductivity of the soil around the cavity could be determined. This method, called the piezometer method, is used for measuring the conductivity of field soil from layer to layer in the soil profile.

If a casing is not driven into a soil hole while it is being augered out, one obtains simply a cylindrical auger hole in the soil and it can be used to measure the hydraulic conductivity of the soil about the hole, provided certain mathematical formulas are found. Kirkham derived the formulas needed in papers published in 1948 and 1958. More detailed theory was later developed in 1971.

To solve a saturated soil-water movement problem, one wishes to know the potential energy in the soil and also the flow paths or streamlines of the water. Kirkham had determined the potential energy and the streamlines for a number of drainage problems.

Streamline functions and potential-energy functions, as Kirkham and his students had so far developed them, could not solve water movement problems where the flow domain was other than a relatively simple shape. Many seepage domains are, however, not of simple shape and to determine flow for them, a whole new mathematical method had to be developed. The needed new mathematical method, which he and his students developed, is called the modified Gram-Schmidt method. Kirkham's expressions are tabulated briefly in a paper published in the *Journal of Geophysical Research* in 1967 with W.L. Powers as senior author. The equations, which give the constants needed for this method of solving problems such as piezometer problems, are listed in more detail in the book *Advanced Soil Physics*, which Kirkham published with W.L. Powers in 1972. Numerous papers involving these functions have been published. Kirkham was an author or co-author in 25 such papers.

Agriculture is charged with polluting rivers and wells through runoff of dissolved chemical fertilizers, pesticides, or animal wastes. A number of pollution problems were solved and the results published by Kirkham and his students. A difficult one concerns the time required for a river that has been polluted to yield polluted seepage flow to wells near the river's side. Others concern the effect of an improper seal around a well and also the effect of a well casing and of a cover slab in stopping pollution. Complicated aquifer shapes and pollution patterns were considered in his publications.

All the foregoing problems have dealt with water-saturated soil. This is a realistic drainage situation because water cannot move from soil into a drainage facility such as a drain tube or a ditch unless the soil at an atmospheric-pressure outflow point is saturated. Nevertheless, water often moves in the soil in the unsaturated state. When Kirkham came to Iowa State University, his first paper was on horizontal unsaturated flow. In this paper, he showed that the wetting front moved by a square root of time law, but that, contrary to the statements of some other workers, this movement was not in accordance with constant-conductivity, heat-flow theory. This paper resulted in British scientists suggesting that a diffusivity or variable conductivity function needed to be used. The idea was followed through by other scientists, and numerous papers on the diffusion theory of capillary flow have subsequently been published.

In irrigated areas, salt moves to the soil surface by unsaturated flow and is removed by saturated leaching flow. In a laboratory experiment, a layer of salt was placed on a porous medium and leaching of the salt into simulated tile drains was performed. The concentration of the drainage water was analyzed as the salt was gradually removed. A theoretical equation was derived for the salt concentration, published in 1972, and it agreed well with the experimental data.

Although Kirkham's papers were mainly on water movement in soil, he contributed to other aspects of soil physics. A paper on sediment transport by rivers and canals gave a theoretical equation that accorded with experimental data. He published papers on heat flow, on air and oxygen movement in soil, and on compaction. He and colleagues developed a neutron or subatomic particle soil probe for measuring soil moisture content. This probe, commonly called the 'neutron probe,' first described in 1952, is still widely used around the world to determine water content of soil. A portable rainfall-simulator infiltrometer was made and patented. Kirkham obtained other patents on soil physics equipment.

The wide application of Kirkham's work is evident from the numerous national and international journals in which his work is published, including the *Soil Science Society of America Journal*; *Water Resources Research*; *Soil Science*; *Transactions of the American Geophysical Union*; *Agronomy Journal*; *Journal of Geophysical Research*; *Transactions of the State Agricultural College, Ghent, Belgium*; *Journal of Irrigation and Drainage of the American Society of Civil Engineers*; *Agricultural Engineering*; *Physical Review*; *Journal of Applied Physics*; *American Math Monthly*; *Journal of Mathematics and Physics*; and many others.

Kirkham's research program attracted and inspired many students to do graduate work at Iowa State University. Under his supervision 72 students were guided towards advanced degrees in soil physics, agricultural engineering, and water resources. Through these students, his contributions to food production and the protection of land against deterioration are being applied in many countries around the world. Without the excellent and large input of capable graduate students into many papers in Kirkham's bibliography, the work could not have been done.

The strong relationships between Kirkham and his students and Kirkham's unique professional contributions are described in the following paragraph quoted from an article written by Dr Dale Swartzendruber, Professor of Agronomy at the University of Nebraska:

> Perusal of Dr Kirkham's research and publications reveals an interesting blend of diversity and specificity. Diversity has been provided by the considerable latitude he has allowed each graduate student in the choice of dissertation research problem. This resulted in problem selections that ranged broadly over the cardinal soil physics topics of soil water (saturated and unsaturated conditions), soil aeration, soil structure and strength, soil temperature and heat, plant–soil–water relationships, and water–solute interactions and processes as related to soil-water and groundwater pollution. Specificity in Dr Kirkham's research program has come from his own special genius – his unparalleled insight, capability, and skill in solving the mathematical boundary-value problems of soil-water movement, particularly as related to saturated flow and drainage. The numerous exact mathematical solutions he has provided establish Dr Kirkham as without peer in this important area of effort. Furthermore, even beyond the solutions as such, there is an additional benefit of great significance. By his direct, explicit, and unapologetic use of the mathematical–physical approach in research, he provided pioneering leadership that cleared the way for others to follow, not only in soil physics, but in related areas as well. For this he might well be called the father of mathematical soil physics.

Further Reading

Kirkham D (1940) Artificial drainage of land: stream-line experiments. The artesian basin. *Transactions of the American Geophysical Union* 21: 587–593.

Kirkham D (1958) Seepage of steady rainfall through soil into drains. *Transactions of the American Geophysical Union* 39: 892–908.

Kirkham D (1967) Explanation of paradoxes in Dupuit-Forchheimer seepage theory. *Water Resources Research* 3: 609–622.

Kirkham D and Powers WL (1972) *Advanced Soil Physics*. New York: Wiley-Interscience.

Kirkham D, van der Ploeg RR, and Horton R (1997) Potential theory for dual-depth subsurface drainage of ponded land. *Water Resources Research* 33: 1643–1654.

Swartzendruber D (1978) In recognition of Don Kirkham on his seventieth birthday. *Soil Science* 125(2): 65–67.

van der Ploeg RR, Horton R, and Kirkham D (1999) Steady flow to drains and wells. In: Skaggs RW and van Schilfgaarde J (eds) *Agricultural Drainage. Agronomy Monograph* 38, pp. 213–263. Madison, WI: ASA, CSSA.

L

Laminar and Turbulent Flow *See* **Hydrodynamics in Soils**

Landfills *See* **Waste Disposal on Land:** Municipal

LAND-USE CLASSIFICATION

J A LaGro Jr, University of Wisconsin-Madison, Madison, WI, USA

Introduction

Land cover is the material cloaking the Earth's surface. Land cover includes water, soil and rock, vegetation, roads, buildings, and other constructed elements supporting human activities. Conversely, land use describes the type and intensity of human activities occurring in a given location. Land-use classification facilitates the monitoring and predicting of urban development patterns, management of water and other natural resources, and land-use planning and regulation.

Land and Society

Spatial variation in climate, physiography, and natural resources has influenced human settlement patterns throughout history. Civilizations have flourished in fertile valleys, along river and lake shores, in coastal areas, and near other highly productive ecosystems. Human population growth has precipitated extensive conversions of undeveloped areas to residential, commercial, and other 'built-up' uses. Factors contributing to population growth and dispersion include the expansion of transportation infrastructure, advances in telecommunication technologies, and a reduction in the differences between urban and rural lifestyles. These are among the many factors that reduce the influence of physiographic constraints on urban development patterns.

Landscapes are transformed when wetlands, prairies, forests, and other habitats are fragmented or replaced by buildings, roads, utility corridors, and other forms of development. These landscape changes impede the movement of wildlife and often occur too rapidly to permit the adaptation of native species through natural selection. Consequently, urban development poses a significant threat to species diversity and ecosystem resilience at both local and regional scales. Urban development may impair the functioning of biophysical systems and reduce these systems' capacity to neutralize pollutants, moderate microclimate, and provide a broad array of other important environmental 'services.' These land-use changes can fragment farms and impair the viability of agriculture on other nearby farms. Development associated with population growth and redistribution may also degrade cultural and historic resources, reduce local outdoor recreational opportunities, and destroy valued views and vistas. The environmental, aesthetic, and economic impacts of these land-use and land-cover changes depend upon the location, or spatial pattern, of development.

Spatial Information and Mapping

A geographic information system (GIS) consists of computer hardware and software, georeferenced

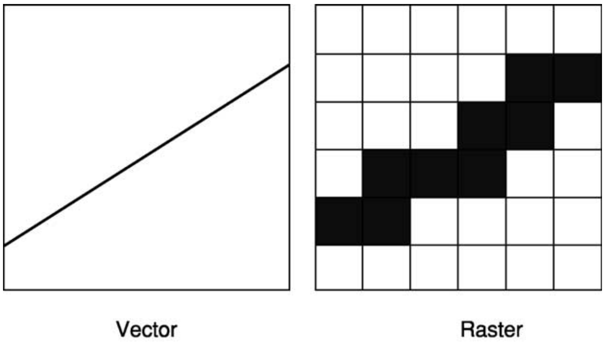


Figure 1 A linear map represented in vector and raster formats.

data on environmental or cultural attributes, and data about the data – or metadata. A GIS can facilitate the organization, analysis, and mapping of land-use and land-cover data. Digital spatial data may be in two forms: raster or vector (Figure 1). Raster data are grid-cell surfaces comprised of hundreds, thousands, or, with large data sets, millions of cells. Vector data, in contrast, consist of arcs, nodes, and polygons. Both types of spatial data are linked to tabular databases that store attribute information about the locations delineated by the grid cells or arcs, nodes, and polygons. The choice of GIS data models depends on the intended purpose of the analysis or map. Land-cover data in a raster format are widely used for environmental modeling and management. Land-use data in a vector format are useful for mapping irregularly shaped cultural features such as roads and land parcel boundaries.

Maps are efficient tools for conveying information about land and landscapes. Maps can show important spatial (or temporal) relationships among the physical, biological, and cultural attributes of a land parcel within its landscape context. Landscape attributes are portrayed with two common types of maps: (1) chloropleth maps; and (2) isopleth maps. A chloropleth map is a thematic map expressing attribute data in discrete categories. Land use, a cultural attribute, can be classified and mapped in a variety of general and more detailed categories. Because this attribute does not imply rank or quantity, land use is expressed on a nominal, or categorical, measurement scale.

An isopleth map, in contrast, displays numerical values of a single attribute. An isopleth map of elevation, for example, partitions into equal increments the entire range of elevations within the mapped area. An isopleth map of elevation might show several elevation classes, each of which represents the site's surface area where elevations fall within equal increments (e.g., 10 m). Each elevation class is simply an aggregated subset of elevations located within the mapped area.

Land-Use Classification

Land-use classification schemes typically address both land use and land cover. A major land-use classification system developed by the United States Geological Survey (USGS) has multiple levels of classification. The categories within these levels are arranged in a nested hierarchy. The most general or aggregated classification (level I) includes broad land-use categories, such as 'agriculture' or 'urban and built-up' land (Table 1). This level of classification is commonly used for regional and other large-scale applications. Within each level I class are a number of more detailed (level II) land-use and land-cover classes. For example, the 'urban and built-up' class includes 'residential,' 'commercial,' and 'industrial' subclasses. Within each of the level II classes, even more detailed classes (levels III and IV) can be defined and mapped. The classes within each level are mutually exclusive and exhaustive. That is, each location within the mapped area can be classified into one and only one class within each level. Together these four levels of classification comprise a hierarchical system for describing, monitoring, and predicting land-use and land-cover change. This standardized, multilevel classification system allows spatially explicit comparisons of land-use inventories conducted over time.

Classifying and mapping land cover is an integral step in understanding the Earth's biophysical systems. Data on the area and distribution of wildlife habitat, for example, are useful in managing and mitigating development impacts on protected and endangered species. Similarly, information on the type, area, and configuration of buildings, roads, and other impervious land-cover facilitates the modeling of storm-water runoff and watershed hydraulics and hydrology.

Different land-cover surfaces reflect sunlight in distinct wavelengths of the electromagnetic spectrum. These spectral 'signatures,' when recorded as remotely sensed images, permit the classification and mapping of individual land-cover types. Sources of data for the classification and mapping of land cover and land use include aerial photographs, satellite imagery, and maps of assorted social and cultural attributes (e.g., property boundaries, population density).

Land-Use Suitability

A location that is suitable for a particular urban land use is one that can accommodate the proposed use with the minimum amount of inputs or resources. This concept of land-use suitability is similar to the United States Natural Resources Conservation Service (NRCS) method for classifying the capability

Table 1 Example of a two-level land-use/land-cover classification system

Level I	*Level II*
1. Urban or built-up	11. Residential
	12. Commercial and services
	13. Industrial
	14. Transportation, communication, and utilities
	15. Industrial and commercial complexes
	16. Mixed urban and built-up land
2. Agriculture	21. Cropland and pasture
	22. Orchards, groves, vineyards, nurseries, and ornamental horticultural areas
	23. Confined feeding operations
3. Rangeland	31. Herbaceous rangeland
	32. Shrub and brush rangeland
	33. Mixed rangeland
4. Forest land	41. Deciduous forest land
	42. Evergreen forest land
	43. Mixed forest land
5. Water	51. Streams and canals
	52. Lakes
	53. Reservoirs
	54. Bays and estuaries
6. Wetlands	61. Forested wetlands
	62. Nonforested wetlands
7. Barren	71. Dry salt flats
	72. Beaches
	73. Sandy areas other than beaches
	74. Bare exposed rock
	75. Strip mines, quarries, and gravel pits

Source: After Anderson JR, Hardy EE, Roach JT, and Witmer E (1976) *A Land Use and Land Cover Classification System for Use with Remote Sensor Data*. Geological Survey Professional Paper 964. Washington, DC: US Government Printing Office.

of soils for agriculture. The NRCS ranks soils by the limitations they present to various types of agriculture. Prime soils, for example, require the fewest inputs for productive agriculture. Locations with these soils do not require irrigation or soil contouring or terracing. These areas also require comparatively smaller amounts of fertilizers and pesticides.

Land-use suitability analysis identifies and maps land areas with different levels of development suitability. The analyst must first identify the attributes that are most relevant to the proposed land uses. The suitability criteria for each land use must then be specified. These suitability criteria include the site conditions that are most desirable, or undesirable, for each land-use type. Suitability criteria for a given land use can vary substantially with differences in local and regional conditions.

The analysis of individual attribute maps is, in many cases, a process of partitioning the spatial distribution of attribute values. A proximity analysis of individual attribute layers may also identify locations within a defined distance of specified points (e.g., entrances), lines (e.g., roads, streams), or areas (e.g., ponds or lakes). A suitability analysis involves three discrete steps:

1. identify suitability criteria for each expected land use;
2. collect and map relevant site attribute data;
3. identify the site locations with attribute values that meet the land-use suitability criteria.

The analysis of an attribute layer may have several objectives, but the primary objective is to find areas that meet a given condition or set of conditions. These areas could be locations on the site where attribute values are:

1. greater than a specified minimum (e.g., elevation of at least 1 m (3 feet) above sea level);
2. less than a specified maximum (e.g., slope gradients less than 20%);
3. within a specified range (e.g., soil drainage is suitable for constructing building foundations and roads).

The analyst can map both favorable and unfavorable conditions for each proposed land use. This suitability map may be the product of either the union or intersection of two or more attribute maps. For example, the intersection of three attribute layers (slope, aspect, and soil drainage) could yield a map showing all locations in an area with slopes less than 20%, a southern aspect, and well-drained soils. This map would have two classes: suitable or unsuitable for the proposed land use.

Subsurface Conditions

Unstable soils can reduce a site's stability for construction of buildings, pedestrian and vehicle circulation systems, and other elements. Other important soil attributes that may be considered, depending on the site's intended uses, include:

1. fertility for plant growth;
2. permeability for drainage;
3. erosion potential;
4. depth to seasonally high water table;
5. depth to bedrock.

The composition and stability of the bedrock that lies beneath, and in some locations at, the Earth's surface are important attributes in land-use planning. The depth to bedrock, in particular, must be assessed if excavations are planned for building foundations or other structures.

When local groundwater is the source of a community's potable water, stringent efforts must be taken to prevent the contamination of this natural resource. Onsite wastewater-treatment systems may utilize soil as the treatment medium. Public health regulations may confine rural residential development to sites that are suitable for soil-based wastewater treatment. Conventional soil-absorption treatment systems function effectively on flat to gently sloping sites with relatively permeable soils containing a combination of sand, silt, and clay. Constraints that preclude the installation of these on-site systems include: high water table, shallow bedrock, steep slopes, flooding, slow substratum permeability, and slow soil permeability.

By increasing the percentage of impervious surfaces in a watershed, land development tends to increase total storm-water runoff volume, reduce groundwater recharge, and increase flooding potential. Increased runoff from land development can increase the risks of flooding, erosion, and habitat degradation. Without mitigation, new construction and other site-disturbing activities compact soils, add impervious land-cover surfaces, and reduce both storm-water infiltration and groundwater recharge. The configuration of impervious surfaces determines their effect on storm-water runoff. Disconnecting impervious surfaces, and separating them with pervious surfaces such as vegetated buffer strips and 'rain gardens,' can help facilitate storm-water infiltration on site. Rain gardens are small, bowl-shaped gardens, planted with perennial native plants and designed to absorb storm-water runoff from roofs, parking lots, and other impervious surfaces.

Poorly draining soils or a shallow water table are development constraints that must be identified early in the land-use planning process. These constraints can be overcome with sump pumps, foundation drains, and other site engineering techniques. But this approach increases development costs and, potentially, exacerbates development impacts on the environment. Sites with a history of prior agricultural use may have experienced significant losses of topsoil. Topsoil depletion increases the costs of re-establishing vegetative cover after construction.

Former commercial and industrial sites should be carefully investigated prior to land-use planning and site development. Gasoline stations, printing presses, dry cleaners, and other commercial sites may be contaminated by a variety of toxic substances. Heavy metals and other toxic substances pose substantial health risks, and if hazardous wastes are present on a site, the costs of remediation can be significant.

Surface Conditions

A topographic survey is essential for most land-use planning projects. Elevation changes create slopes that have both a gradient and an orientation, or aspect (the compass direction that the slope faces). A site's suitability for roads, walkways, buildings, and other structures is partially a function of the existing slope gradients on the site. Climate also affects the influence of slope on site suitability. The maximum slopes at which roads and walkways are safe for transportation are not the same in a climate with freezing temperatures and a climate where temperatures are always above freezing. Particularly in locations with seasonal freezing temperatures, pavement gradients must be relatively low (typically less than 12%) to prevent slipping on icy surfaces. Variations in slope gradient and aspect affect the levels of solar radiation striking the site's surface or objects on the site. This can influence the amount of energy expended in heating and cooling buildings on the site. Elevation changes also affect drainage patterns, which influence plant distributions.

Hurricanes, earthquakes, and other natural hazards pose significant risks to human health, safety, and welfare. Often the loss of life and property can be prevented, or at least mitigated, by reducing these risks. Yet vulnerability to natural hazards is not uniformly distributed across the landscape. Sites on the edges of water may be attractive properties for real-estate development. If the land is inherently unsuited for development, however, homes and other structures in these locations can be at serious risk. Rainfall associated with hurricanes and other long-lasting storms typically saturates soils, fills up surface water-storage capacity, and increases the risk of local flooding. The main threat to life and property is the storm surge (the water pushed toward the shore by the force of the storm's winds); therefore, development in low-lying areas is especially vulnerable.

Preventing development of naturally hazardous sites, rather than attempting to control nature with technology, is both economically efficient and environmentally sound. Confining land development to sites without significant constraints reduces construction costs, protects inherent natural amenities, and allows the continuing function of critical environmental systems.

Land-Use Planning

Land-use changes are a direct result of the technologies, institutions, and values of human society. Critical environmental resources can be protected by directing growth and development to more suitable areas served by public infrastructure. At the community or municipal scale, land-use planning determines the locations of new roads, recreation areas, commercial, industrial, and residential developments, and other land uses. Municipal land-use planning is one way government protects public health, safety, and welfare. Land-use planning within the public sector is increasingly important in protecting environmental quality.

At the parcel or site scale, land-use planning determines the locations of individual buildings, pedestrian and vehicle circulation systems, and other built elements. Site-scale planning is initiated by individuals, corporations, and by agencies or departments of local, state, and federal governments. Land-use regulations, local real-estate markets, and cultural and physiographic attributes all influence, to varying degrees, a site's capacity to accommodate a particular land-use program. Zoning ordinances and other local codes may dictate permitted land uses, building heights and densities, and the location of buildings on the site. Federal and state regulations may also affect a myriad of land-use planning decisions.

Site Selection

Property value is affected by the size and shape of the parcel. Larger sites can accommodate more development and offer greater flexibility in the types and spatial arrangement of uses on the site. Property appraisers in the USA attempt to estimate value based on the property's 'highest and best use.' For residential and some commercial uses, scenic views to lakes, rivers, or other environmental features may contribute significantly to market demand and property value. Parks and other recreational facilities are amenities that also enhance property value. Long and linear properties have a much higher edge-to-interior ratio than properties that are more square in shape. Greater 'exposure' to the surrounding landscape may reduce the site's ability to buffer negative off-site impacts. However, if the adjacent areas are natural or cultural amenities, a parcel with a relatively high edge-to-interior ratio can be desirable, particularly if these amenities are in public ownership and are likely to remain so well into the near future.

Site-selection decisions involve consideration of economic, social, environmental, and even ethical, factors. Potential sites are evaluated and selected, or rejected, on the basis of each site's suitability for the intended land uses (i.e., program). Rugged topography, unstable soils, and shallow bedrock are just a few of the constraints that may deter development. Development on difficult sites, particularly if the site constraints are not avoided or the impacts are not mitigated, can have serious on-site and off-site consequences. These constraints also increase design and construction costs substantially.

Typically, one or two factors have a disproportionately large influence on the site-selection decision. These factors may be essential site attributes that, if absent, exclude a site from further consideration. For example, the primary site-selection criterion for a development of single-family homes in an non-sewered area may be the suitability of the soil for on-site wastewater-treatment systems. If the essential criteria are met, the site-evaluation process can proceed to consider other selection criteria. Severe constraints, however, would exclude the site from further consideration.

Programming

A land-development program is simply a description of the intended site uses. The programming phase often involves visiting the site, inspecting aerial photographs, performing market analyses or user-demand studies, and reviewing other available site and user information. The site's suitability for future development or restoration depends, however, on the land-use program under consideration. Site attributes relevant to a proposed nature conservation area, for example, would differ significantly from the site attributes relevant to a proposed housing development. A nature conservation program would place a high priority on a diversity of wildlife habitats, e.g., from dry upland south-facing slopes to lowland marshes and bogs. In contrast, the housing project might place a high priority on gently sloping areas without development constraints like shallow bedrock, shallow water table, or protected endangered species.

Site Inventory

The site inventory is an important step in understanding the physical, biological, and cultural linkages

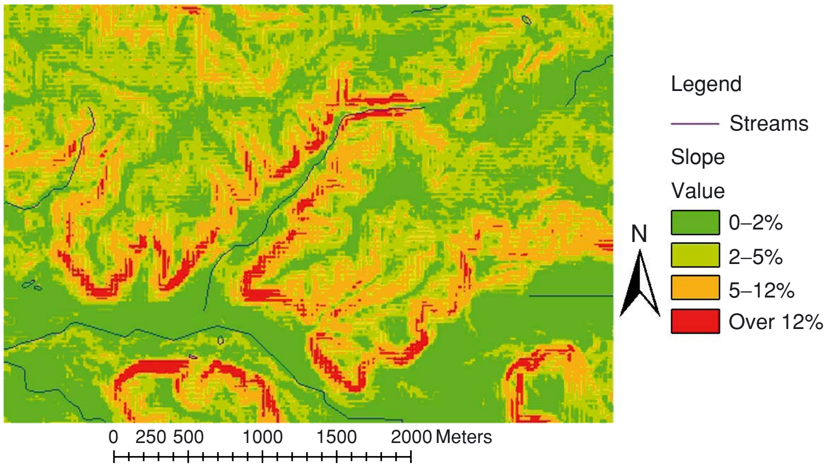


Figure 2 A topographic slope gradient map, showing four classes of slope gradients.

between the site and the surrounding landscape. The inventory is not, however, an open-ended investigation of biophysical or cultural phenomena. The project program, in conjunction with the site's physical, biological, and cultural features, determines which data should be collected in the site inventory. Consider, for example, a partially wooded and hilly site along a small, but navigable river. Several site attributes could be mapped, but all of these may not be relevant to any given program. A map showing the site's range of slope percentages (e.g., 0–2%, 5–12%) (Figure 2) could be potentially useful because some land uses are better suited for gentle slopes, and other uses are better suited for steeper slopes.

Sources of data for the site inventory include topographic surveys, aerial photographs, soils maps, and other maps created to support the land-use planning process. Ecologists, hydrologists, anthropologists, and other experts may participate in collecting, mapping, and analyzing the site-attribute data. The inventory documents the key biophysical and cultural attributes of the site and adjacent land. An inventory map of existing vegetation, for example, shows the existing site conditions for a single attribute – vegetation type. This map will remain valid whether the proposed uses of the site include intensive development or no development at all. Maps of site drainage patterns, steep slopes, or other site attributes are similarly unaffected by changes in the proposed land-use program.

Site Analysis

The site analysis is a program-driven assessment of a site's physical, biological, and cultural attributes. This analysis is tailored to each unique land-use program. Attribute maps in the site inventory provide the data needed for this program-specific analysis. The analysis of site attributes is made tractable by considering the assets and liabilities, or opportunities and constraints, that the site poses for the proposed land-use program. Information contained in physical, biological, and cultural attribute maps can be synthesized to create maps of the site's suitability for each expected land use (Figure 3). The liabilities (constraints) associated with a site and its context reduce the site's suitability for one or more of the proposed uses. Conversely, the assets (opportunities) associated with a site and its context increase the site's suitability for the proposed uses. Site amenities are features that have social, economic, ecological, or aesthetic value. Amenities warranting protection include features that enhance the visual quality of the site and the surrounding community. These include specimen trees, natural drainage swales and depressions, unique rock outcrops, scenic overlooks, and historic or culturally significant buildings and structures.

Plan Development and Implementation

Land-use suitability is a function of the site's soil conditions, slopes, microclimate, and other biophysical and cultural attributes. Land-use plans locate proposed activities, infrastructure, and other physical improvements on the site. Implementing a plan might include the construction of buildings, roads, walkways, or various other site structures. Site constraints increase the development's complexity and cost. By utilizing modern engineering technology, some land uses can be accommodated in locations that are inherently unsuited for those uses. For example, housing could be built in marshy areas, on highly erodible soils, and on a wide variety of other 'difficult' sites. Preparing these areas for development might require tree removal, extensive regrading and drainage, and modification of natural storm-water runoff and infiltration

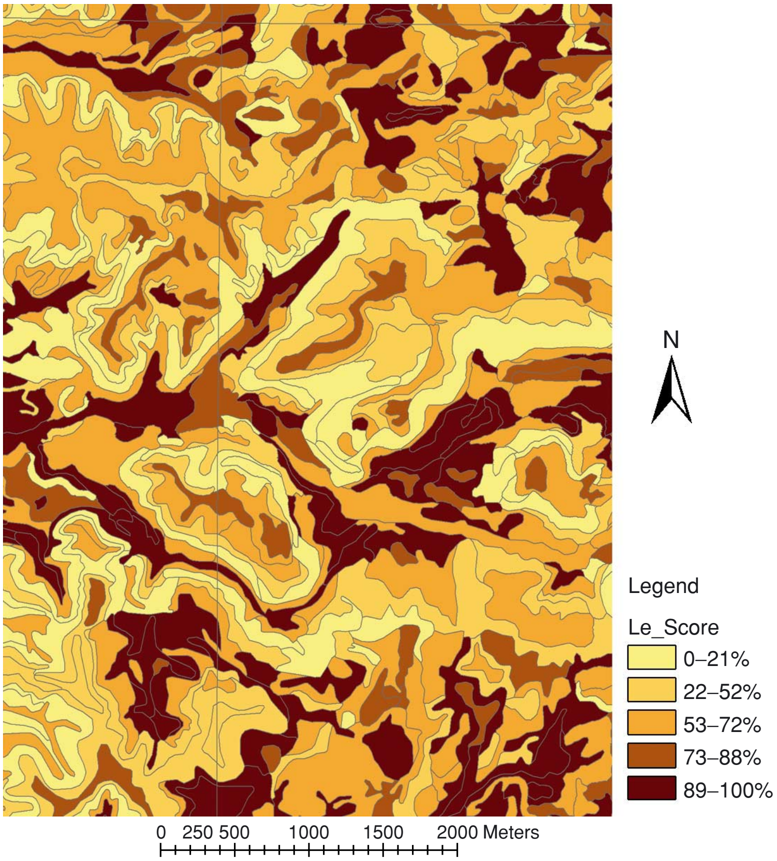


Figure 3 A land evaluation map showing a range of site suitability for residential development; higher scores indicate greater development suitability.

patterns. Yet development in unsuitable areas can destroy natural aesthetic amenities, degrade unique wildlife habitat, and alter long-standing hydrological regimes. This 'brute force' approach can also have significant negative environmental impacts.

Conclusion

Low-impact land-use planning protects the land's ecological 'infrastructure' from disturbances associated with development. This ecologically sensitive approach to planning can also involve environmental restoration. For example, restoring a riparian community along an eroded stream bank can stabilize soils, filter sediments and chemical pollutants from storm water, and create habitat for birds, insects, and other beneficial species. These land-planning decisions require accurate and spatially explicit information on the land's suitability for proposed uses. With accurate and comprehensive knowledge of the area's opportunities and constraints, land planners can mitigate, and often prevent, environmental impacts associated with population growth and redistribution.

See also: **Classification of Soils**; **Geographical Information Systems**

Further Reading

Anderson JR, Hardy EE, Roach JT, and Witmer RE (1976) *A Land Use and Land Cover Classification System for Use with Remote Sensor Data*. Geological Survey Professional Paper 964. Washington, DC: US Government Printing Office.

Beatley T and Manning K (1997) *The Ecology of Place: Planning for Environment, Economy, and Community*. Washington, DC: Island Press.

Forman RTT and Godron M (1986) *Landscape Ecology*. New York: John Wiley.

Goudie A (1994) *The Human Impact on the Natural Environment*, 4th edn. Cambridge, Massachusetts: The MIT Press.

LaGro JA Jr (2001) *Site Analysis: Linking Program and Concept in Land Planning and Design*. New York: John Wiley.

Lynch K and Hack G (1984) *Site Planning*. Cambridge, Massachusetts: The MIT Press.

McDonnell MJ and Pickett STA (eds) (1993) *Humans as Components of Ecosystems*. New York: Springer-Verlag.

McHarg IL (1969) *Design with Nature*. Published for the Natural History Press. Philadelphia: Falcon Press.
Mileti DS (1999) *Disasters by Design: A Reassessment of Natural Hazards in the United States*. Washington, DC: Joseph Henry Press.
Odum EP (1989) *Ecology and Our Endangered Life-Support Systems*. Sunderland, Massachusetts: Sinauer Associates.
Zonneveld IS and Forman RTT (eds) (1990) *Changing Landscapes: An Ecological Perspective*. New York: Springer-Verlag.

LAWES, JOHN BENNET AND GILBERT, JOSEPH HENRY

A E Johnston, Rothamsted Research, Harpenden, UK

In agricultural science the names of Lawes and Gilbert are forever linked, and together they are regarded as the founding fathers of the scientific method in agricultural research. Gilbert took up his appointment with Lawes at Rothamsted on 1 June 1843, the date now regarded as the start of Rothamsted Experimental Station (now called Rothamsted Research). This was the beginning of one of the longest (57 years) and most productive scientific partnerships on record. Their skills and characters were totally complementary. Lawes, the landowner and practicing farmer, was versatile and farseeing, quick to appreciate the needs of the farmer and able to respond with sound practical and economic advice based on the results of the Rothamsted experiments. But the quality of these results owed much to Gilbert. Everything he undertook was performed with scrupulous accuracy and meticulous care to detail. Gilbert indeed gave scientific expression to the work of Rothamsted and established the experiments on farm animals and crops on a sound scientific basis. Together Lawes and Gilbert published some 150 scientific papers and 300 or so popular articles for farmers (Figure 1).

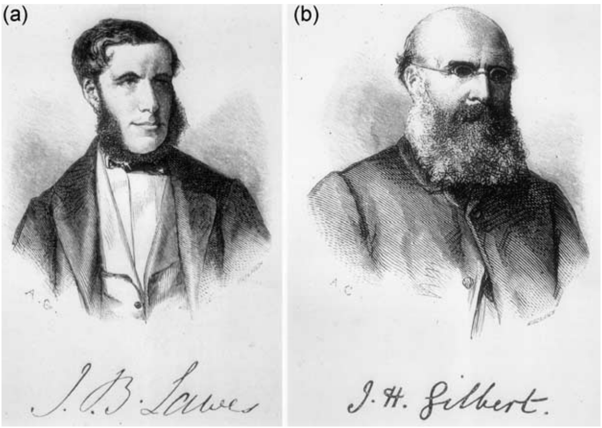


Figure 1 Lawes and Gilbert as young men: (a) J. B. Lawes; (b) J. H. Gilbert. Photographs probably taken in the early 1850s.

It is impossible to conceive that, in 1843, Lawes and Gilbert could have had the remotest idea that some of the field experiments they started on agricultural crops during the next 13 years would still be continued in the twenty-first century. Today these experiments are the oldest, continuous field experiments in the world and their importance is recognized worldwide. With modifications introduced only after very careful consideration, they still provide scientific data of immeasurable value not only to farming practice but also to the wider issues of the sustainability of farming systems and the interaction of agriculture and the environment. It is worth considering, therefore, the background to the start of these experiments and of their founding fathers.

John Bennet Lawes

Lawes (later Sir John Lawes) was born in 1814 at Rothamsted, a small agricultural estate near Harpenden, some 40 km north of London. When his father died in 1822, his widowed mother with her two children, John and Caroline, temporarily had to leave the manor house for financial reasons. During this period, John went first to Eton and then, as an undergraduate, to Oxford. When he was 20, Lawes' mother returned to live at the manor house. So, before taking his degree, Lawes decided to leave Oxford to manage the Rothamsted estate of approximately 100 ha on behalf of the family. He appears to have had few regrets at leaving university, because he found little to stimulate an inquisitive scientific mind in the formal teaching offered at that time.

While at Oxford, however, Lawes must have had contact with Professor C.G. Daubeny, FRS, whose experiments and writings were probably a major formative influence on Lawes. Daubeny simultaneously held the Chairs of Chemistry, Botany, and

Rural Economy and his research and writings from the 1820s onward ended a very barren period in the development of agricultural science in the UK. Two of Daubeny's interests are important to subsequent developments at Rothamsted. He was doing experiments in the Botanic Garden on the effect of a crop on the growth of a subsequent crop (allelopathy) and writing about the essential need for good experiments to help understand and improve agricultural practices.

Joseph Henry Gilbert

Born in 1817 and therefore very close in age to Lawes, Gilbert's background was very different to that of Lawes. As a boy, Gilbert (later Sir Henry Gilbert) had, according to his mother, only three interests: electricity, arithmetic, and chemistry. He studied chemistry at university, first at Glasgow then at University College London. After graduating he worked initially with Professor Anthony Todd Thompson, founder of the British Pharmacopoeia. Early in 1840 he went to work with Professor Justus von Liebig at Giessen, in Germany, and was awarded a doctorate within 12 months; apparently the criteria were different then from now! Returning to England, he was at University College London in 1841, before going to an industrial appointment in Manchester. Lawes may have met Gilbert when the latter was in London. Lawes had at least visited Thompson's laboratory and that of Thomas Graham, Professor of Chemistry and a world authority on phosphorus chemistry at that time. Later, in 1843, Graham, who was very friendly with Liebig, wrote to him asking him to receive Lawes at Giessen, but there is no record of a visit taking place.

The Agricultural Background in the Second Quarter of the Nineteenth Century

Lawes' and Gilbert's achievements need to be set against farming practice and the state of knowledge in the 1830s. Agronomic techniques had been changing only slowly during the previous centuries. Food production in the middle ages was based on the open field system. Although the land in each village belonged to a single landowner, the arable land was divided into two or three large fields and most villagers were allowed to cultivate small, scattered strips of land in each field. Each field was cropped in rotation, usually fallow followed by a cereal, or fallow followed first by a legume and then a cereal. Each villager aimed to produce at least enough grain to feed his family. In addition each villager's animals had access to common grazing land. Landlords started to create distinct fields, by enclosing, with hedges or fences, open grazing land in the fifteenth and sixteenth centuries. Later, in the latter half of the eighteenth and early nineteenth centuries, similar enclosure of the open arable fields occurred. Thus, landlords divided their land into discrete farms, each with part arable crops and part permanent grass. Although these enclosures allowed the opportunity for greater efficiency in the use of land, they created a large population who had no access to land to produce their own food.

This increased efficiency was aided by two inventions by Jethro Tull in the mid eighteenth century. The first was the seed drill and second, the horse-drawn hoe. This allowed seed to be placed at a uniform depth, improving germination, in regularly spaced rows, and the interrow space to be kept relatively free of weeds by mechanical hoeing. There then developed a four-course rotation of arable crops: roots, spring barley, a legume (either herbage, e.g., clover, or a grain legume, e.g., peas or beans) and winter wheat (the Norfolk four-course rotation). The introduction of the root crop (swedes, turnips, mangels) and the herbage legume allowed more animals to be kept over winter, with some return of plant nutrients via the farmyard manure (FYM) produced. However, supplies of FYM were limited and that which was available was usually applied to the root crop. Winter wheat, producing the staple food of the people, was grown only once in 4 years and must have obtained its nitrogen (N) from the residues of the leguminous crop that preceded it. This rotation became the backbone of arable cropping first in England and then in much of temperate Europe. Landlords insisted that their tenants followed it because it minimized nutrient losses from the farm and the build-up of pests and diseases, and it helped control weeds. There was one problem, however: the root crop was sown usually in early June and if it did not germinate and grow quickly the emerging seedlings could be killed by the turnip flea beetle. This decreased yields and the amount of food available for animals in the winter.

In the early years of the nineteenth century, there were still major gaps in knowledge of crop nutrition. Following the studies of de Saussure and others, it was generally accepted that plants acquired their mineral nutrients (phosphorus, P; potassium, K; calcium, Ca; magnesium, Mg; etc.) from the soil, but there was considerable uncertainty about the source of nitrogen (N). From analyses done in Alsace, France, it was known that FYM contained both minerals and nitrogen (N).

In Britain at about this time, N was becoming available as very impure ammonium sulfate, a by-product of the manufacture of town gas, and as sodium nitrate, imported from Chile. Potassium salts were available in limited supply, mainly from

Russia and North America. They were obtained by burning plant material (mainly trees) in large metal pots and extracting the soluble salts from the ash with water, hence the name 'potash,' the word frequently used for potassium in agriculture. The large natural deposits of K salts in Germany were not discovered until the 1860s.

In Britain in the early 1800s, it was observed that bone trimmings, from the manufacture of bone handles for steel cutlery, when applied to land speeded the germination of turnips and hence the yield. This led to a considerable trade in bones that was specific to the UK. During 1837–42 the estimated mean annual British domestic supply was 27 000 t, while imports averaged approximately 46 000 t. Justus von Liebig accused the British of sustaining these imports by robbing the battlefields of Europe and he maintained that the imports amounted to the manurial equivalent of an army of some 3.5 million men!

The Years at Rothamsted Immediately Before 1843

Returning to Rothamsted in 1834 to manage the estate, Lawes had a bedroom at the manor house converted to a laboratory and attempted to determine the 'active principles' of the many medicinal plants, including poppies, he started to grow on the farm. With only one untrained person to help, he was unsuccessful. However, a useful lesson was probably learnt – the need for skilled help and rigorous analytical methods. He also attempted to develop commercially a newly proposed method for the manufacture of calomel, mercurous chloride, then used as a purgative in medicine. Again he was without success, but he must have learnt some valuable lessons about 'trade' (running a commercial enterprise), not then normally a part of a Victorian country gentleman's education and training, but soon to be put to good use by him.

The Earliest Experiments

In 1837 and 1838, Lawes did some simple experiments using cabbages and turnips, both used for animal feed, and grown either in pots or on small plots on the farm fields. He compared a range of nitrogen-containing salts, each applied at the same rate of salt. He was clearly puzzled by the fact that the yields were not increased in proportion to the amount of N applied, but he did comment: "phosphate of ammonia to be one of the most powerful manures known." In a second article in the *Gardeners' Chronicle*, Lawes noted that applying ammonia to a soil deficient in P was useless and, where minerals were in adequate supply, the ammonia supplied by the atmosphere would be insufficient for the wants of a crop.

Early Observations on the Need for Phosphorus and the Production of Superphosphate

Once he was managing the farm, Lawes noted, as did other farmers in the locality, that crushed bones, seen to be so efficacious in certain parts of the country, did not increase the yield of his or his neighbors' turnips. Like others, however, Lawes found that if bones were treated with sulfuric acid to produce superphosphate of lime then, when this was applied to land, the yields of all the arable crops grown on the farm were greatly increased. Lawes did many experiments to determine the ratio of bones to acid and the acid strength to produce a product that was dry but hard so that it could be crushed to a fine powder and be spread by machine. In 1842, Lawes took out a patent for his process and besides bones he included mineral phosphates as a source of phosphorus. In 1843 he started production of superphosphate on a commercial scale at a factory at Deptford, London. This was the start of what was to become a successful fertilizer business. A much larger factory was opened at Barking Creek in 1852.

It was from the profits of his business enterprises (he also had a factory producing citric acid) that Lawes financed the agricultural experiments at Rothamsted. No doubt bearing in mind the pleas of Daubeny for well-conducted agricultural experiments, Lawes wisely decided that he could not be both a fledgling industrialist with a business in London and an aspiring agricultural researcher at Rothamsted 40 km away. In consequence Lawes appointed Gilbert to help set up and supervise the experiments on the nutrition of both animals and crops that he was keen to start.

The Start of the Experimental Program at Rothamsted

In many parts of the world today, Rothamsted is associated with those long-term, 'Classical' field experiments on arable crops and permanent grass started by Lawes and Gilbert between 1843 and 1856 that still continue (Table 1). Lawes, however,

Table 1 The principal field experiments started by Lawes and Gilbert at Rothamsted and the crops each investigated

Field	*Year experiment started*	*Crop*
Barnfield	1843	Turnips
Broadbalk	1843	Winter wheat
Geescroft	1847	Beans
Hoosfield	1849	Clover
Hoosfield	1852	Spring barley
Geescroft	1869	Oats
Agdell	1848	Crops in rotation
Park Grass	1856	Permanent pasture

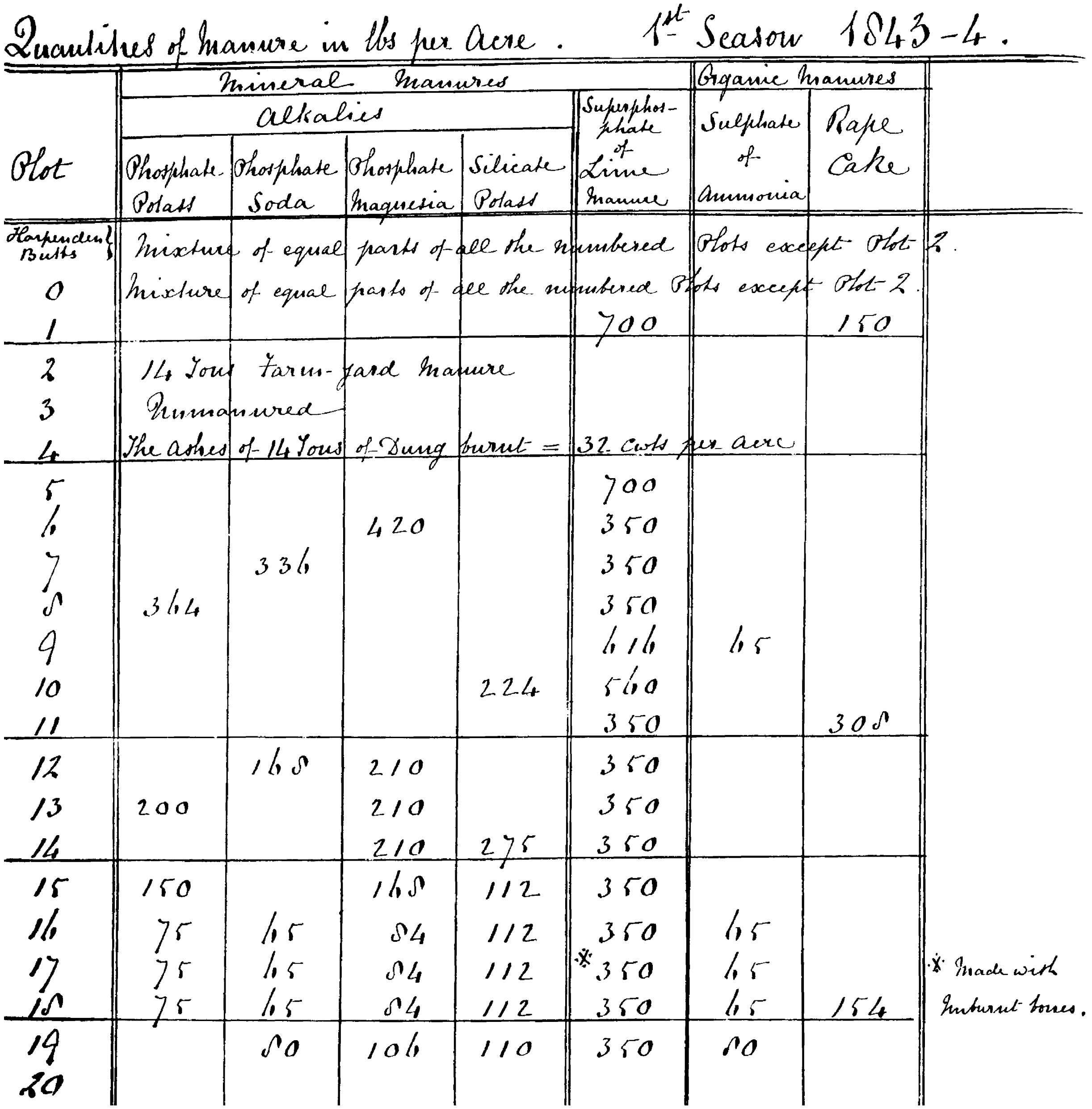
was also interested in animal nutrition, and the pioneering experiments undertaken at Rothamsted laid the sure foundation for this branch of agricultural science. In fact, in the early years of the Rothamsted experiments, more effort was probably expended on these animal studies than on those with crops. Research on animals gradually diminished and ended in the 1890s, after Lawes and Gilbert had published some 50 scientific papers and articles on animal nutrition.

With regard to the field experiments on crops, Lawes had made some important decisions about their conduct before he appointed Gilbert. Probably the most important was to have separate experiments for each of the principal arable crops grown in rotation. This decision was probably a direct result of his contact with Daubeny whilst at Oxford and the results from the latter's studies of the effect of one crop on a subsequent crop. Over the period 1843–69, Lawes and Gilbert started an experiment on each of the arable crops grown most frequently in the Norfolk four-course rotation and, in 1856, an experiment on the nutrient requirements of permanent grassland (Table 1). Lawes and Gilbert quickly learnt that running a large-scale field experiment program not only required their skills in planning, designing, and supervising, but also, and equally important, the farm staff had to acquire the appropriate working skills for the field operations (Figure 2). The decision to grow each crop in monoculture illustrates Lawes' scientific approach, because the decision flew in the face of the accepted agricultural practice of crop rotation. Lawes justified the decision on the grounds that this not only would highlight the characteristic nutrient requirements of each crop, but also would emphasize differences between crops.

Another important decision taken at the start of the experiments identifies Lawes as a farmer. He decided to have large plots, believing that fellow farmers were more likely to accept the results if they could be seen to be consistent over a large area. Initially on Broadbalk there were 21 long, narrow plots (each about 322 × 7.5 m) on approximately 4.4 ha, the width being related to the 'lands' into which the field was divided for cultivation (Figure 3). These large plots were but one among the many legacies left us by Lawes, because it has been possible to subdivide the plots to test changes in treatment. Lawes must also have decided the treatments to be tested on the Barnfield experiment on root crops, because it was started very soon after Gilbert came to Rothamsted on 1 June 1843. Besides an unmanured control and FYM, the other treatments were probably based not only on his experience, but also on the views expressed by Liebig in 1840. In his now-famous report, commissioned by the British Association for the Advancement of Science, entitled *Organic Chemistry in its Application to Agriculture and Physiology*, which was presented in 1840 and also first published in that year, Liebig stressed the need to apply the mineral elements, i.e., plant nutrient elements other than nitrogen. The availability of simple, soluble inorganic salts of many of these mineral elements and of nitrogen made it possible for Lawes and Gilbert to test them singly and in various combinations, and compare the yields with those given by FYM as well as the unmanured control. The treatments initially tested on the winter wheat experiment started on Broadbalk in autumn 1843 (Figure 2) were very similar to those on Barnfield.

In the first year of the Barnfield experiment, only eight of 23 plots on Barnfield tested nitrogen and then only up to 15 kg N ha^{-1}. The effects of N were far from clear. The method of applying fertilizer and drilling the seed brought both into close proximity and germination was impaired, especially with ammonium salts. Where plants survived they were individually larger than with minerals only, but the increase in total yield was small because plant numbers were less. Similarly for the first crop of winter wheat sown in autumn 1843 on Broadbalk, nitrogen was tested on only five of the 21 plots. But in sharp contrast to the variable effects of N on Barnfield, winter wheat grain yields at harvest in 1844 were much increased by the modest amount (15 kg ha^{-1}) of nitrogen tested as an ammonium salt. Compared with the unmanured crop, minerals alone increased yield by 10%, minerals plus N by 35%. Lawes had not used cereals in his earlier experiments, because it was not normal practice to apply nutrients to wheat that followed a legume in the Norfolk four-course rotation. Whether or not this response to N was a surprise is unknown, but Lawes and Gilbert were quick to appreciate the significance of the result. For the next crop of winter wheat, drilled in autumn 1844, N was tested on 14 of the 21 plots and at four rates, 12, 24, 36, and 48 kg N ha^{-1}.

Lawes and Gilbert went on to prove that it was more important for crops, other than legumes, to have an adequate supply of readily available N than of minerals. Because soils at that time were usually deficient in P, yields with N, P, and K were always largest, but the effect of N was greater than that of P and K (Table 2). The experiments also showed that inorganic fertilizers applied in appropriate combinations and amounts could give the same yield as the large amount (35 t ha^{-1}) of FYM tested (Table 3). Lawes and Gilbert never said that fertilizers were better than FYM. What they foresaw was that fertilizers would have to be used to supplement the limited supplies of FYM if there was to be sufficient food



Quantities of Manure in lbs per Acre. 1st Season 1843-4.

Plot	Mineral Manures					Organic Manures	
	Alkalies						
	Phosphate Potass	Phosphate Soda	Phosphate Magnesia	Silicate Potass	Superphosphate of Lime Manure	Sulphate of Ammonia	Rape Cake
Harpenden Butts	Mixture of equal parts of all the numbered Plots except Plot 2.						
0	Mixture of equal parts of all the numbered Plots except Plot 2.						
1					700		150
2	14 Tons Farm-yard Manure						
3	Unmanured						
4	The ashes of 14 Tons of Dung burnt = 32 Cwts per acre						
5					700		
6			420		350		
7		336			350		
8	364				350		
9					616	65	
10				224	560		
11					350		308
12		168	210		350		
13	200		210		350		
14			210	275	350		
15	150		168	112	350		
16	75	65	84	112	350	65	
17	75	65	84	112	*350	65	
18	75	65	84	112	350	65	154
19		80	106	110	350	80	
20							

* Made with unburnt bones.

Figure 2 The manure sheet for the first crop of winter wheat sown autumn 1843 on Broadbalk and harvested in 1844.

for the rapidly expanding urban population. Their results also allowed them to say with some confidence that yields could be maintained by using fertilizers. Figure 4 shows that, over the subsequent years, yields of winter wheat on Broadbalk have not only been maintained but have increased where fertilizers are used. Similarly, yields on the other classical experiments have been increased as new, higher-yielding cultivars have been introduced. Clearly in these experiments the soil, climate, crop, and management have been in 'harmony'; excellent examples of the sustainability of crop production under these conditions.

In the second half of the nineteenth century, the demand for N fertilizers increased to such an extent that in 1898, Sir William Crookes, in his presidential address to the British Association, warned that the world's known reserves of combined N were being rapidly exhausted. Fortunately it was not long before scientists and chemical engineers discovered how to fix atmospheric nitrogen. This was done first by the Norwegians Birkeland and Eyde, who combined nitrogen and oxygen to produce nitrate in an energy-intensive process. Later the more effective process of combining nitrogen with hydrogen to form ammonium, the Haber–Bosch process, was developed.

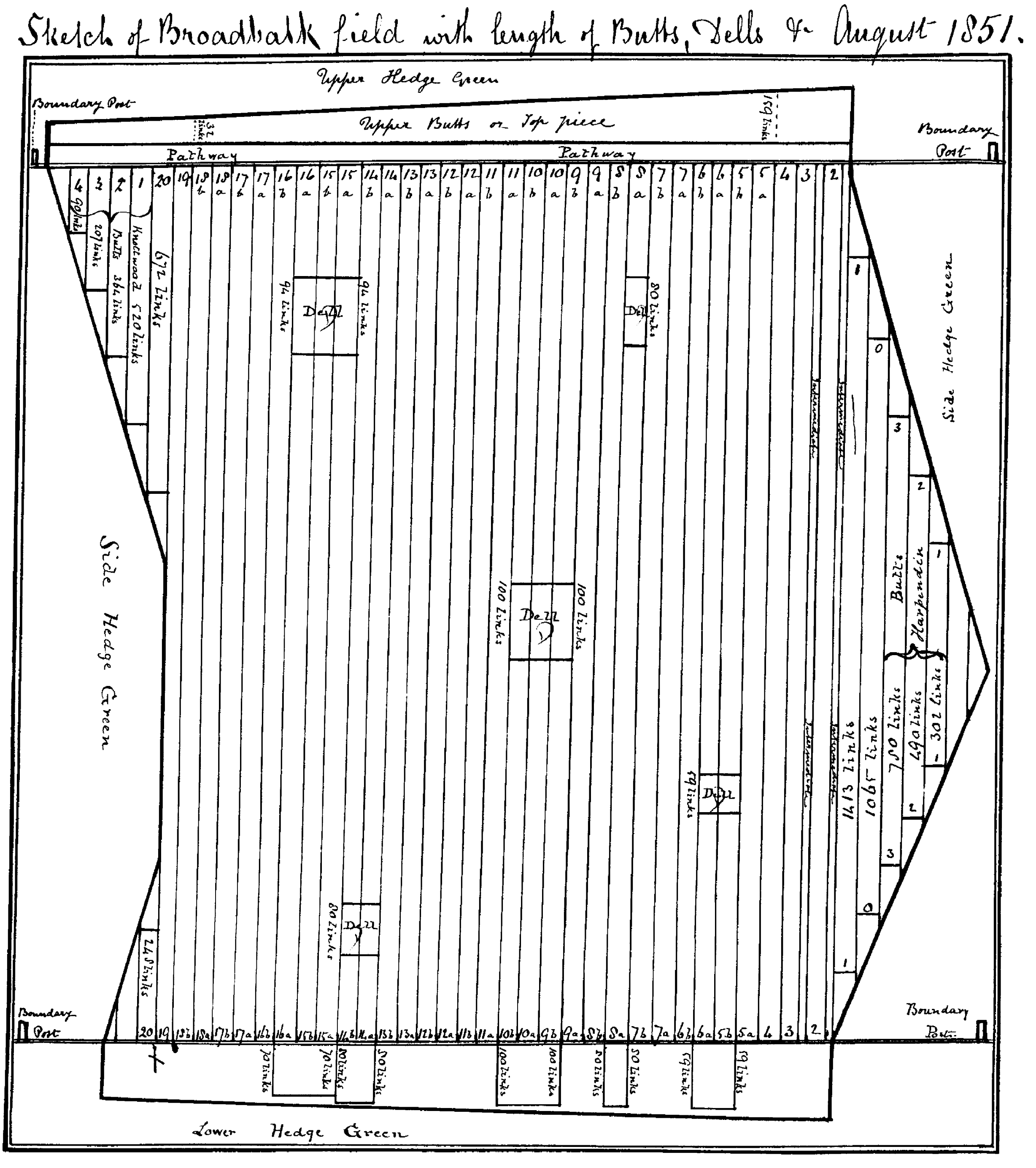


Figure 3 Division of the Broadbalk field at Rothamsted, 1851.

As they began to appreciate the questions to ask of their experiments, Lawes and Gilbert modified the treatments tested in the early years on Barnfield and especially on Broadbalk. The changes on the latter were designed to ascertain the direct and residual effect of fertilizers supplying ammonium and nitrate nitrogen, the major topic of debate with Justus von Liebig in Germany. They also decided that, although large plots were important, long narrow plots, as on Broadbalk and Barnfield, were not the best arrangement. When the experiment on spring barley was started on Hoosfield in 1852, the plots were squarer (42 × 17.5 m) than those on Broadbalk, and the treatments were arranged so that P and K were tested on strips running east–west, which were crossed at right angles with strips testing N. Outside of the main block of 16 plots (nil, P, K, PK × nil and three sources of N, ammonium and nitrate salts, and rape cake, an

Table 2 Yields (tonnes per hectare) of winter wheat, spring barley, mangels, and beans in the early years of the field experiments at Rothamsted

Experiment field name	*Crop*	*Years*	*Yield (t ha^{-1})*			
			None	*PK*	*N*	*NPK*
Broadbalk	Winter wheat, grain	1852–61	1.12	1.29	1.63	2.52
Hoosfield	Spring barley, grain	1872–81	0.85	1.10	1.66	2.62
Barnfield	Mangels roots	1876–84	9.5	11.3	25.6	46.0
Geescroft	Field beans[a], grain	1847–62	1.11	1.50	NT	1.64

[a] *Vicia faba.*

NT, not tested.

Table 3 Yields (tonnes per hectare) of winter wheat and spring barley, grain at 85% dry matter, and roots of mangels and sugar beet at Rothamsted

Experiment field name	*Crop*	*Period*	*Yield (t ha^{-1})*	
			FYM[a]	*NPK fertilizers*[b]
Broadbalk	Winter wheat	1852–56	2.41	2.52
		1902–11	2.62	2.76
		1955–64	2.97	2.85
		1970–75	5.80	5.47
Hoosfield	Spring barley	1856–61	2.85	2.91
		1902–11	2.96	2.52
		1952–61	3.51	2.50
		1964–67 (PA)[c]	4.60	3.36
		1964–67 (MB)[c]	5.00	5.00
Barnfield	Mangels[d]	1876–94	42.2	46.0
		1941–59	22.3	36.2
	Sugar beet	1946–59	15.6	20.1

[a] FYM 35 t ha^{-1} containing on average 225 kg N.

[b] N fertilizer per hectare; barley, 48 kg; wheat, 144 kg; root crops, 96 kg.

[c] PA, Plumage Archer (cultivar) given 48 kg N ha^{-1}; MB, Maris Badger (cultivar) given 96 kg N ha^{-1}.

[d] *Beta vulgaris.*

organic N source), there were initially three plots: a second unmanured plot, one testing FYM, and the third testing the dried, powered soil ash added to the weighed amount of fertilizer for each plot so that there was a constant volume of material to be spread. A further improvement in experimental design was made in 1856 on Barnfield. Three of the original long, narrow plots were combined and assigned to test FYM, P, and K in various combinations, and these strip treatments were crossed at right angles by the same nitrogen treatments tested on Hoosfield. Thus only on Barnfield was fertilizer nitrogen applied with FYM, and this invariably gave the largest yields of mangels. The plot arrangement was not changed on Broadbalk, but there were various changes to the treatments before 1852, and from then little further change until recent times.

The Place of Long-Term Experiments in Agricultural Research and the Role of Lawes and Gilbert

Long-term or continuing experiments are the best practical way of assessing the sustainability of an agricultural system. Comparing different husbandry systems often allows the opportunity to explain failures if these occur and allows modifications to be introduced to maintain, and where appropriate, increase productivity. In this respect it is important to ask why Lawes and Gilbert continued with some of their experiments.

In part, the initial impetus to continue may have come from their various controversies with Liebig. Probably the most important argument was about the source of nitrogen for plants. In the first, 1840 edition, of his book, Liebig wrote: “Cultivated plants receive the same quantity of nitrogen from the atmosphere as trees, shrubs and other wild plants; but this is not sufficient for the purposes of agriculture.” In 1843 Liebig changed the last phrase, and the sentence reads: “Cultivated plants receive the same quantity of nitrogen from the atmosphere as trees, shrubs and other wild plants; and this is quite sufficient for the purposes of agriculture.” Results from the field experiments at Rothamsted supported the 1840 statement, not that of 1843. Yet for some 20 years Liebig stubbornly refused to retract his 1843 statement, and an acrimonious debate was continued in the pages of various journals.

Other arguments involved Liebig’s frequently repeated statement that the cropping and manuring at Rothamsted would rapidly exhaust the soil of minerals and so diminish its productivity. This did not happen. Another source of disagreement was Liebig’s assertion that it was only necessary to replace the quantity of minerals removed in the harvested crop. The Rothamsted experiments showed that this was not true; to get good yields, it was essential to apply more minerals than the crop contained. This was probably because at that time the soils contained little

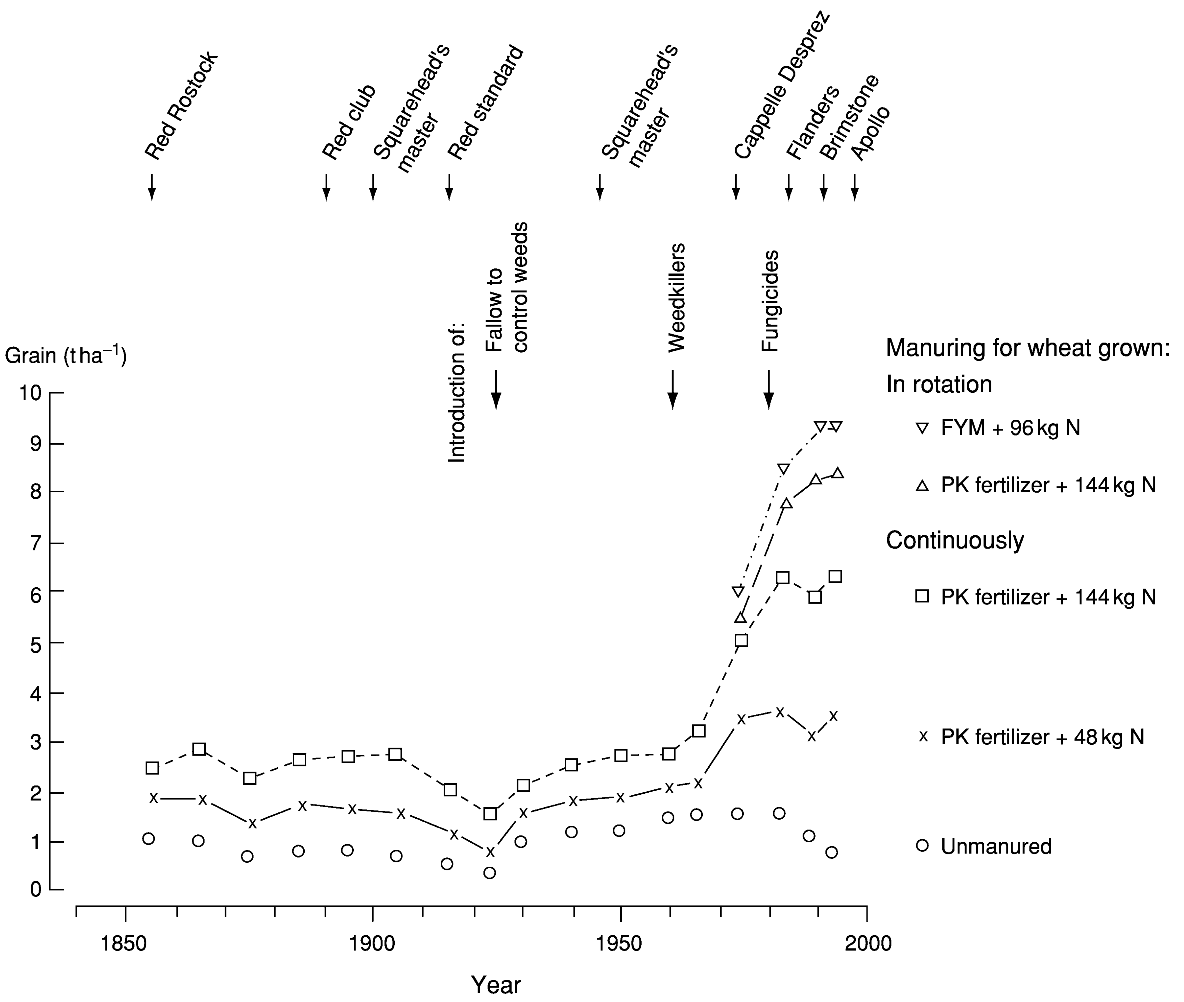


Figure 4 Yields of winter wheat cultivars (tonnes of grain per hectare) from the Broadbalk experiment, Rothamsted Research, Harpenden, UK.

plant-available P and K. Lawes and Gilbert may have needed to go on showing that they were not only right, but also right beyond all reasonable doubt. And so they kept the experiments going, lest Liebig claimed that the experiments were stopped because he was right and Lawes and Gilbert were wrong. Perhaps the benefits to agricultural science from the Rothamsted experiments owe something to Liebig! A lesson for today from these controversies may be that healthy competition, springing from a lack of consensus, may lead to greater and more rapid progress than complacency resulting from consensus!

Lawes was a great educationalist, using the results from the large-scale field experiments to offer sound advice to farmers. On 19 July 1855, in his response to the Earl of Chichester on the occasion of the presentation to him of the newly built Testimonial Laboratory, funded by public subscription, Lawes said "the object of these investigations (at Rothamsted) is not exactly to put money into my pocket but to give you the knowledge by which you may be able to put money into yours." Sir John Lawes did make a lot of money from his fertilizer business, as well as licencing others to make superphosphate using his process. He could readily have become a business tycoon, owning and living on a pleasant country estate, but his intense interest in agriculture decreed otherwise. As the field experiments continued, there was ever-greater opportunity both to 'fine tune' advice given previously and seek new interpretations of the accumulating data. In consequence, Lawes came out strongly on the importance of continuous investigations in agriculture and later, when summarizing the results of the first 50 years, Lawes and Gilbert said that the experiments had become more valuable with time.

Lawes maintained financially the Rothamsted experiments from his own resources and, using a third of the proceeds from the sale of the fertilizer business in 1872, he endowed trustees in 1889 to ensure that the experiments he had started with Gilbert might continue. In 1935, money raised by public subscription allowed the trustees to purchase the estate when Lawes' descendants offered it for sale. The purchase ensured the security of the land with its world-renowned classical and other long-term field experiments. Today financial support comes from other sources, but agricultural and ecologic sciences, and the interface of both with the wider environment, reap great benefit from Lawes' generosity and Lawes and Gilbert's scientific objectivity and foresight. Rothamsted now has a well-documented database of results from these field experiments, extending over a period of more than 150 years, and a unique archive of crop and soil samples collected over the same period. The archives can be and have been mined for answers to questions Lawes and Gilbert could not have conceived. But besides the past there is the future: six of the experiments started between 1843 and 1856 still continue, all with carefully considered modifications. These experiments should have at least as much to offer in the future as they have yielded in the past.

Further Reading

Dyke GV (1980) *John Lawes of Rothamsted – Pioneer of Science, Farming and Industry*. Harpenden, UK: Hoos Press.

Dyke GV (1991) *John Bennet Lawes: The Record of His Genius*. Chichester, UK: John Wiley.

Johnston AE (1990) Liebig and the Rothamsted experiments. In: Judel GK and Winnewisser M (eds) *Symposium "150 Jahre Agrikulturchemie"*, pp. 37–64. Germany: Justus Liebig-Gesellschaft zu Giessen.

Johnston AE (1991) *Potential Changes in Soil Fertility From Arable Farming Including Organic Systems*, pp. 1–38. Proceedings No. 306. York, UK: International Fertiliser Society.

Johnston AE (1994) The Rothamsted Classical Experiments. In: Leigh RA and Johnston AE (eds) *Long-term Experiments in Agricultural and Ecological Sciences*, pp. 9–37. Wallingford, UK: CAB International.

Johnston AE (1997) The value of long-term experiments in agricultural, ecological and environmental research. *Advances in Agronomy* 59: 291–333.

Johnston AE and Powlson DS (1994) The setting up, conduct and applicability of long-term, continuing field experiments in agricultural research. In: Greenland DJ and Szabolcs I (eds) *Soil Resilience and Sustainable Land Use*, pp. 395–421. Wallingford, UK: CAB International.

Lawes JB (1842) Ammoniacal manure, p. 221. *The Gardeners Chronicle*. London, UK: Bradbury and Evans.

Lawes JB (1843) Ammonia, p. 692. *The Gardeners Chronicle*, London, UK: Bradbury and Evans.

Lawes JB (1882) The future of agricultural field experiments. *Agricultural Students Gazette, New Series* 1: 33–35.

Lawes JB and Gilbert JH (1895) The Rothamsted experiments over fifty years. *Transactions of the Highland and Agricultural Society of Scotland. Fifth Series* 7: 11–354.

Leigh RA and Johnston AE (eds) (1994) *Long-term Experiments in Agricultural and Ecological Sciences*. Wallingford, UK: CAB International.

Liebig J von (1840) *Organic Chemistry in its Application to Agriculture and Physiology*. London, UK: Taylor Walton.

Orwin CS (1938) *The Open Fields*. Oxford, UK: Clarendon Press.

Orwin CS (1949) *A History of English Farming*. London, UK: Thomas Nelson.

Pirie NW (1991) The role of Rothamsted in making nutrition a science. *The Biochemist* 13: 8–13.

Powlson DS and Johnston AE (1994) Long-term field experiments: their importance in understanding sustainable land use. In: Greenland DJ and Szabolcs I (eds) *Soil Resilience and Sustainable Land Use*, pp. 367–394. Wallingford, UK: CAB International.

Thompson FML (1968) The Second Agricultural Revolution 1815–1880. *Economic History Review,* 2nd Series. XXI: 62–77.

LEACHING PROCESSES

B E Clothier and S Green, HortResearch, Palmerston North, New Zealand

Introduction

Soil is the thin and biologically active skin of our "Green Planet." Soil, by virtue of its intrinsic fertility, plays a key role in sustaining our primary-production systems. To enhance the productivity of these agricultural systems, irrigation water, fertilizers, and pesticides are often applied to crops growing on soil. Also, soils are sometimes used for the disposal, either intentional or accidental, of certain waste products; and the soil surface is the depository of atmospheric compounds that might be either natural, such as salt, or compounds from human activities, such as sulfur. Since soil lies astride the main pathway along which rainwater leaches to groundwater or surface water bodies, the water quality of these receiving water reserves is strongly influenced by land management, and the myriad of interlinked water-transport processes and chemical-exchange mechanisms that operate in the soil domain. Development of sustainable land-management strategies, and the design of environmental-protection protocols, demand that the leaching process be well understood.

Leaching encompasses the gamut of biophysical mechanisms and chemical exchange processes that can result in altered chemical concentrations in any pore water that is percolating through soil. There are two aspects to leaching: the quantity of the water that is moving through the soil and destined for subterranean or surface water bodies, and the chemical concentration, or quality, of that pore water. We use the term 'leaching' to encompass two meanings: the deep seepage of water, and the export of chemicals carried by that seepage.

There are controls on the quantity of drainage water quitting the base of the root zone (*D*, in millimeters). For risk assessment it is necessary to know both the concentration of a chemical in the leachate (*C*, in moles per cubic meter) and the total chemical load (per unit ground area) that is moving to the bodies that receive water (*M*, in moles per square meter; or, more commonly, kilograms per hectare). The total leaching of chemical, *M* (load), is the product of *D* (quantity) and *C* (quality). Thus, in the discussion of leaching processes, the transport of water and the chemical processes operating in soil are both considered.

Physically, soil is a porous medium comprising a complex arrangement of three phases: solid, liquid, and air. Much to the chagrin of scientists, convective and diffusive flow through soil does not behave in the uniform and isotropic way assumed by their simple theoretical descriptions. Rather, aggregates, cracks, and macropores serve to create an apparently chaotic flow regime that leads to rapid and far-reaching transport of chemicals. Modeling methods account for this preferential flow and are utilized for risk-assessment purposes.

Leachate Quantity

Globally, an average of some 720 mm of rainfall (RF, in millimeters) falls on soil each year. The many nooks and crannies in the soil provide it with a capillary attractiveness so that there is some tendency for the infiltrating rainwater to be absorbed by the soil, near the surface. The ever-present force of gravity tends to draw complementary water deeper into the profile beyond the grasp of the roots. On balance, approximately 410 mm of the global average rainfall is lost back as vapor to the atmosphere, either by direct free-water evaporation from the soil and directly off plant surfaces (*E*, in millimeters) or by transpiration through plants (*T*, in millimeters). Any transpired water must first have been captured from the soil by the plant's roots. The complement of evaporation, namely drainage, *D*, therefore averages approximately 310 mm per year, because, on an annual timescale, the change in soil-water storage is negligible. Drainage water acts as a vehicle, carrying with it passenger solutes that can leach the soil of chemicals that may compromise the quality of receiving waters. For productive systems, rainfall might be augmented by irrigation (*I*). Indeed, irrigation comprises approximately 80–85% of the world's consumption of fresh water.

Both for production purposes and environmental protection, it is critical to determine whether the partitioning of root-zone water and its dissolved solutes is either into the plants by uptake or despatch by drainage (*D*) toward groundwater. Roots are the arbiters of this partitioning, and so they determine water and solute fate (Figure 1). In simple terms, roots penetrate the soil to an effective depth, z_R (millimeters). From within this root zone they extract liquid water and some dissolved solutes. Thus rhizosphere processes critically affect both the quantity and quality of leachates leaving the root zone.

The water balance of the root zone, over a given time interval Δt, can be written to find the drainage at depth z_R as:

$$D = \mathrm{RF} + I - R - E - T - \Delta S \qquad [1]$$

where ΔS is the change, over Δt, of the storage of soil water in the root zone (millimeters). The variation in

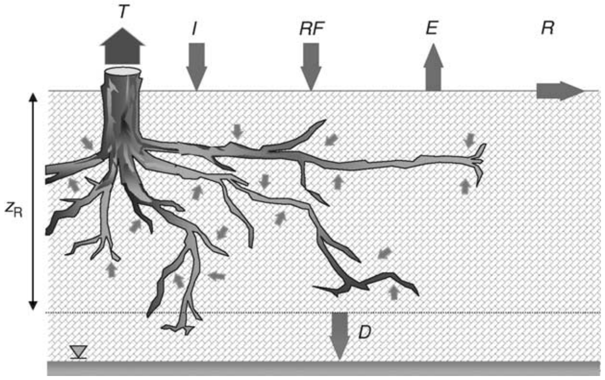


Figure 1 The water balance of the root zone. The effective depth of rooting is z_R. Inputs into the root zone include rainfall (RF), as well as irrigation (*I*) in some cases. Not all of the applied water might infiltrate, as there could be runoff (*R*). Evaporative losses are either by direct evaporation from the surface (*E*), or by transpiration from plants (*T*). Drainage of water from the base of the root zone toward groundwater, *D*, represents the quantity of leaching.

the storage of water can be found from the change in the soil-water content profile, $\theta(z)$ (cubic meters per cubic meter), over Δt, integrated from the surface down to the base of the root zone:

$$\Delta S = \int_0^{z_R} \Delta\theta(z)\mathrm{d}z \qquad [2]$$

Thus the quantity of leaching, D, is implicitly linked to the weather (RF, E, and T), as well as to the nature of the soil and the vegetation cover, plus the effective depth and distribution of plant roots.

At a mechanistic level, this interplay between soil-water diffusion and convection, and root-water extraction, can be described by the Darcy–Buckingham equation:

$$\frac{\partial\theta}{\partial t} = \frac{\mathrm{d}\theta}{\mathrm{d}h}\frac{\partial h}{\partial t} = \frac{\partial}{\partial z}\left(K(h)\frac{\partial h}{\partial z}\right) - \frac{\partial K(h)}{\partial z} - U(h(z), t) \qquad [3]$$

where h (in millimeters) is the soil-water pressure head, which is a function of soil's water content, and K (millimeters per second) is the pressure head-dependent, soil-water conductivity function. The first term on the right-hand side describes the diffusive nature of the capillary-induced movement of water through soil. The second term accounts for direct convection in response to gravity. The last term, U (millimeters per second), is the depth-dependent extraction of water by plant roots. Many forms have been proposed for U, and all rely on the theoretical insight provided by Gardner on the uptake of water by roots.

The Darcy–Buckingham equation is the theoretical scheme used to model and predict the partitioning of soil-water fate into either plant uptake (T), or drainage (D).

Leachate Quality

The chemical concentration of the soil's pore-water solution, C, depends on a complex of linked interactions between transient transport processes and chemical exchange mechanisms, as well as production, extraction, and degradation reactions. For any chemical species, the convection–dispersion equation describes the resident concentration at any time and location in the soil as:

$$\frac{\partial\theta C}{\partial t} + \frac{\partial\rho S}{\partial t} = \frac{\partial}{\partial z}\left(\theta D_s\frac{\partial C}{\partial z}\right) - \frac{\partial q_w C}{\partial z} + \sum\Gamma_i \qquad [4]$$

where S is the concentration of chemical that is retained by the soil matrix (moles per kilogram), ρ is the soil's bulk density (kilograms per cubic meter), D_s is the solute's diffusion–dispersion coefficient (millimeters squared per second), q_w is the water flux density (millimeters per second), and Γ_i is the source/sink term for the i reactions that can produce, extract, or degrade the chemical species concerned (moles per cubic meter per second). The structure of this formula (eqn [4]) highlights the inextricable link between water movement (q_w; eqn [3]), and chemical-transport processes.

Also, eqn [4] accounts for the link between the chemical concentration in the soil's solution (C) and the amount of chemical that is adsorbed to the surfaces of the soil's matrix (S). This relationship is called the adsorption isotherm. One commonly used form that describes this exchange process is the Langmuir isotherm:

$$S = S_{\max}\frac{mC}{1 + mC} \qquad [5]$$

where $S_{\max}$ is the maximum adsorption (moles per kilogram) and m is an empirical constant (cubic meters per mole). Often, however, for simplicity, or in an absence of better information, a linear isotherm is assumed:

$$S = K_D C \qquad [6]$$

where K_D is called the distribution coefficient (cubic meters per kilogram).

Equations [3] and [4] provide a mathematical framework that is commonly used to interpret and predict the leaching processes that determine the quantity of leachate (D) and the leachate load of chemical (M).

Preferential Leaching Processes

The macroscopic theory on which eqns [3] and [4] is based requires that the soil be a uniform medium. Unfortunately, soil is a highly structured medium, such that the three phases of solid, liquid, and air are not isotropic. Indeed there is often a high degree of connectedness between these phases. Soil structure is typically characterized by aggregates, fractures, and biogenic macropores (**Figure 2**).

The spatial topology of the macroporous networks that typify structured soil determines the speed and extent of preferential flow processes. As a result of local nonequilibrium processes, there can be more rapid and deeper-than-expected leaching of surface-applied chemicals, relative to that predicted by eqns [3] and [4].

Not all of the soil's pore water would therefore seem to be actively involved in water transport and chemical exchange. Aggregates and macropores can

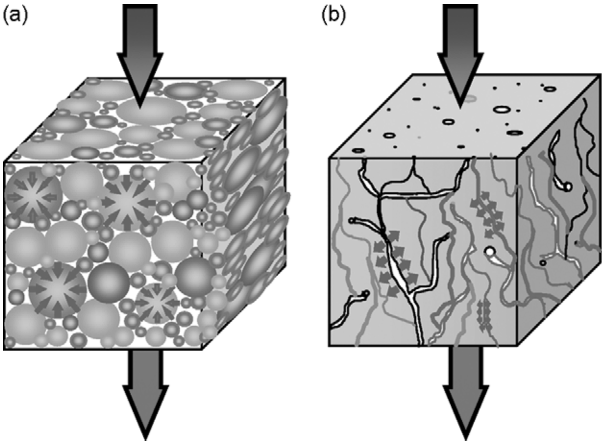


Figure 2 Soil structure is characterized either by aggregates (a) or macropores and fractures (b). Such structure results in rapid and far-reaching transport processes that lead to preferential leaching of applied chemicals. In such structured soils, there is meanwhile likely to be reduced leaching of chemicals produced from within the soil's matrix.

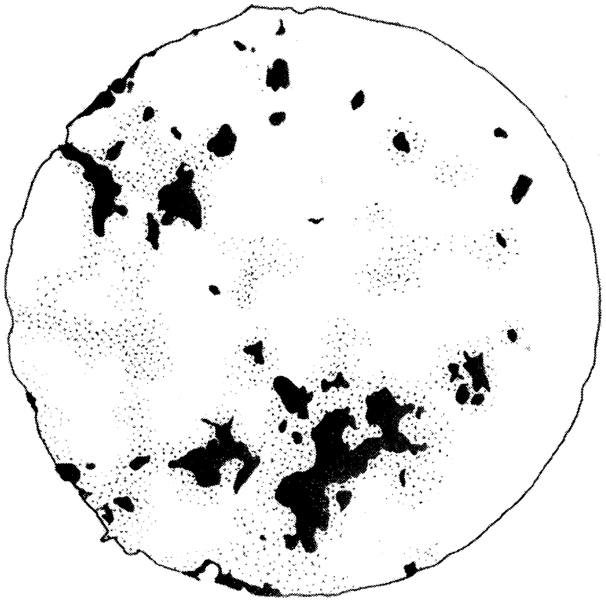


Figure 3 Contrasting patterns of the leaching of two dyes through a 120-mm-diameter core of undisturbed soil. The dark area was stained by a dye tracer contained in saturated flow from a free-water pond at the surface ($h_o = 0$), and the stippled area was from dye leached by unsaturated water applied at the slight negative head of $h_o = -20$ mm. (Reproduced from Scotter DR and Kanchanasut P (1981) Anion movement in a soil under pasture. *Australian Journal of Soil Research* 19: 299–307, with permission of CSIRO Publishing.)

create a wide spectrum of pore-water velocities associated with a solution that is either invading the soil or leaching from it. In an aggregated soil, the more-rapid water-flow paths will be those that are connected throughout the interaggregate domain. So interaggregate water will preferentially wend its way around the aggregates, which tend to possess smaller, albeit connected intra-aggregate pores. These would not seem to be as effective at leaching chemicals from root-zone soil, because the flow velocities are lower and the connected path-lengths shorter.

Alternative, nonmechanistic schemes have been developed for modeling the preferential leaching of chemicals through structured soil. Process-based transport schemes are difficult to parameterize, and so transfer-function models have been considered, because they are less data-intensive and they can support either a deterministic or a stochastic framework.

Pedogenic and biogenic processes can impart upon soil a macroporous structure characterized by a longitudinal connectedness. The enhanced mobility of flow down these variously sized macropores, as predicted by Poiseuille's law of flow through a pipe, is abetted by macropore connectivity to produce preferential transport of the invading solution or enhanced leaching of some part of the resident solution.

In an illustrative study of preferential leaching in the laboratory with undisturbed cores of soil that was growing pasture, two dye tracers were used to mimic leaching. Methylene blue dye, applied in free water ponded on the soil's surface ($h_o = 0$), was found to stain only a small volumetric fraction of the soil's volume (Figure 3). When water containing rhodamine-B dye was applied to the soil in water at the slight unsaturated surface potential of -20 mm, the pattern of leaching was found to involve a larger volumetric fraction of the soil's porosity. Obviously, leaching in structured soils can, depending on hydraulic conditions, involve differing volumes of the soil's wetted porosity.

In one of the most popular mechanistic models used currently to model the preferential leaching that results from the presence of either aggregates or macropores, Occam's razor is simply used to cut the soil's volumetric water content θ (cubic meters per cubic meter) in two. This binary assumption does not account for changes in the leaching processes as a function of the soil's degree of unsaturation (cf. Figure 3), but it does better-represent the preferential leaching processes that prevail in field soils. The great simplicity of this model is that one fraction of the soil's water, θ_{im} (cubic meters per cubic meter), is considered, for leaching purposes, to be stagnant. The complementary portion of the soil's water, θ_m (cubic meters per cubic meter), is deemed to be mobile, as might be imagined of the water that is between the aggregates or flowing in the macropores.

A two-domain approximation for the leaching of a conservative solute under steady flow at pore-water velocity v_m, (millimeters per second) is given by:

$$\theta_m \frac{\partial C_m}{\partial t} + \theta_{im} \frac{\partial C_{im}}{\partial t} = \theta_m D_s \frac{\partial^2 C_m}{\partial z^2} - \nu_m \theta_m \frac{\partial C_m}{\partial z} \qquad [7]$$

where C_m and C_{im} are the concentrations of inert chemical in the mobile and immobile domains, and $\nu_m = q_w/\theta_m$. It is assumed that molecular diffusion in the direction of leaching in the mobile domain is negligible, and that the only mechanism to transfer solute between domains is, for analytical convenience, limited to a diffusion-like process that can be described by:

$$\theta_{im} \frac{\partial C_{im}}{\partial t} = \alpha(C_m - C_{im}) \qquad [8]$$

where α (per second) is a first-order mass transfer coefficient.

In order to measure θ_{im} independently, a tracer-filled disc permeameter is used to allow leaching into the soil of a solution at a concentration C_m. Ignoring any effects that might accrue from interdomain diffusion (eqn [8]), the flux concentration in the mobile domain in the soil just under the disc is C_m. Thus by removing soil samples from under the disc and determining the average resident concentration of solute C^*, the mobile fraction can be found using the partitioning formula:

$$\theta C^* = \theta_m C_m + \theta_{im} C_{im} \qquad [9]$$

If a tracer not initially present in the soil is used and if diffusive exchange (eqn [8]) is considered sufficiently weak, then C_{im} remains essentially zero over the time scale of the experiment. So:

$$\theta_m = \theta\left(\frac{C^*}{C_m}\right) \qquad [10]$$

via which θ_m can be obtained independently from measurements immediately under the disc of just C^* and θ. The results for a silt loam suggest that, for leaching, θ_m/θ is approximately 0.7–0.75 at a pressure head of $h_o = -40$ mm. Only about three-quarters of the soil's water-filled porosity appears to be actively involved in leaching. This qualitatively corroborates the cross-sectional staining areas in **Figure 3**, where it can be presumed that approximately 50% of the area is pore space, and of that only about 75% is stained by dye.

A novel aspect has been added to this technique through the use of the sequential invasion of i inert tracers into the same soil under the permeameter. They show how differences in the respective resident concentrations of C_i^* reflect the rate of diffusion, α, into the immobile domain. The resident concentration in the soil of any tracer i reflects the diffusion that has taken place during the time t_i it has been present in the mobile domain. From a regression of the measurements using:

$$\ln\left[1 - \frac{C_i^*}{C_{m,i}}\right] = \frac{-\alpha}{\theta_{im}} t_i + \ln\left[\frac{\theta_{im}}{\theta}\right] \qquad [11]$$

both θ_m and α can be deduced. From many experiments on various soils, α has been found to range between approximately 5×10^{-7} and 5×10^{-6} s^{-1}.

For aggregated soils, there is a numerical model, the so-called MACRO, which allows convection between the interaggregate mobile zone and the intra-aggregate pores. This convection is considered gravity-free and is driven simply by water content differences between the respective domains. Solute transfer between these domains is allowed by diffusion, as well as by direct convection with the transferring water. By simulation, "… at short times, macropore flow increases leaching of non-reactive solute. At longer times, leaching is reduced by macropore flow [and it is] critical to account for the effects of macropore flow in clay soils on the leaching of sorbing, degrading compounds."

Modeling and Risk Assessment

So, despite the complicated mathematical forms of the coupled flow and transport eqns [3] and [4], and even [7] with [8], modern computers and efficient software schemes are easily able to solve them to predict the impact of land management on leaching processes. The results from models can then be used in a risk-assessment framework to help us better understand the functioning of environmental systems and to guide decisions for sustainable land management.

A simple, lumped-parameter model of the leaching processes under short pasture that is grazed and a pine forest provides insight into the processes of leaching in terms of the controls on leachate quantity (D), and the contrast between the chemical concentration in the leachate (C) and total leachate load (M). The simulation considers pasture and pines growing on soil typical of the Nelson region of New Zealand. Actual weather data from 1972 to 1998 are used. The maximum rooting depth, z_R, of 0.5 m is assumed for pasture, whereas the pine trees are considered able to root effectively to any prescribed depth. Simulations using a lumped-parameter model have been carried out for $z_R = 0.5$, 1, 1.5, 2, 2.5, and 3 m (**Figure 4**).

When pasture and pines are grown on soils of a similar shallow depth of 0.5 m, the evapotranspiration (ET) for the pines is lower due to sparseness of the canopy in the early stages of tree growth and again following the thinnings. None the less, due to

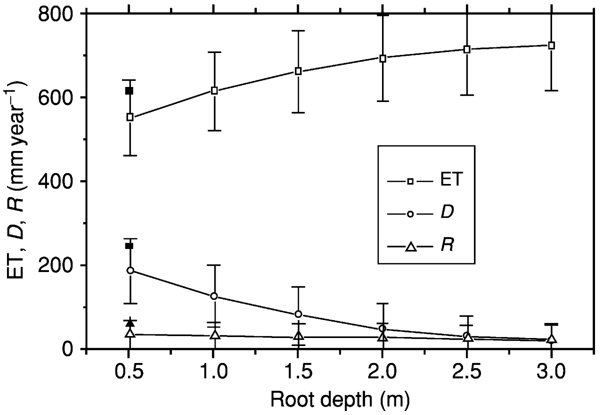


Figure 4 The influence of the depth of the root zone on the amount of leaching (*D*) 'lost' by pine trees (open symbols) in comparison with pasture (closed symbols). Pasture is assumed to root to a maximum depth of z_R of 0.5 m, whereas the pines can extend their roots beyond 3.0 m if soil depth permits. The mean annual rainfall over the 27 years was 974 mm. Reproduced with permission from Clothier BE, *et al. Modelling Rootzone Processes: a Tool for Risk Assessment*, Water 2000, NZWWA.

interception losses by the trees reducing the effective rainfall, there is still a lower leachate quantity (*D*) under pines for such shallow soils. As the soil depth increases, the rise in ET constricts leachate quantity (*D*) under the pines. For a shallow soil of $z_R = 0.5$ m, the annual leachate load under pines is approximately 195 mm, or 20% of RF. When the soil is deep, for example $z_R = 3$ m, $D = 23$ mm or just 2.5% of RF. Thus the amount of water leaching through soil is critically dependent on the nature of the vegetation cover and soil's properties, and in particular the depth of effective rooting z_R.

Pasture production and pine tree growth are modeled through a nitrogen-balance scheme. Consequently, this model can be used to assess the risk posed to groundwater quality by the leaching of nitrate from the root zone of these crops. Figure 5 shows the average annual nitrate-leaching concentrations, *C*, and the total annual loadings, *M*, for the pasture ($z_R = 0.5$ m) and the pines with $z_R = 2$ m.

There are two distinctive characteristics: the difference between the two land uses, and the distinctly skewed and non-Gaussian nature of the cumulative probability functions (CPFs). The average concentration (probability 0.5) is much lower for pines (2.3 mg l^{-1}) than for pasture (12.1 mg l^{-1}), since a lot more of the nitrogen ends up stored in the plant tissue. Through repeated grazings, and additions of dung and urine, there are greater inputs of N into the pastoral system. The CPF for the trees is the most non-Gaussian, and this is due to the infrequent high loadings of nitrate that occur when, occasionally, a wet year follows a dry year and there is a surge in nitrate leaching. The simple, lumped-parameter model provides the amount of leaching (*D*) so that this can be multiplied by the nitrate concentration (*C*) to obtain a loading of nitrate of the receiving waters (*M*). The model results in Figure 5 show that the loading on groundwater from grazed pasture is predicted to be 41 kg · N ha^{-1} per year. However for the pine trees growing on this 2-m-deep soil, the low concentrations of nitrate combined with the low amount of drainage mean that there is simulated to be a very low leachate loading of just 0.6 kg · N ha^{-1} per year. So whereas pine trees are predicted to have a negative impact of the quantity of leaching (*D*), they are considered, by virtue of a lower total loading (*M*), to have a beneficial role in maintaining the quality of leachates moving to receiving bodies. The interplay between land management and soil character determines the quantity and quality of leaching.

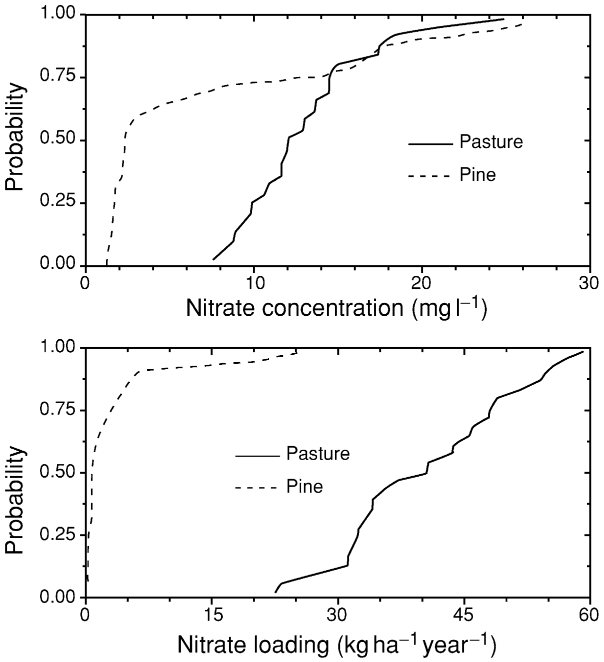


Figure 5 The predicted cumulative probability functions (CPF) of: (a) the concentration of nitrate in the leachate under pasture, and that under pine trees (soil depth 2 m; but the rooting depth for pasture is $z_R = 0.5$ m and pines, $z_R = 2.0$ m). This CPF relates to the annualized value that was derived by running the model in daily time-steps over 27 years, and summing the results to form annual totals; and (b) the total loading of nitrate on the receiving water bodies. Reproduced with permission from Clothier BE, *et al. Modelling Rootzone Processes: a Tool for Risk Assessment*, Water 2000, NZWWA.

A more-sophisticated modeling framework, the Soil–Plant–Atmosphere System Model (SPASMO), predicts the risks posed by the leaching of the active ingredient (AI) of three pesticides applied, in a simulation, to a vineyard on Twyford silt loam located above an unconfined aquifer in Hawkes Bay, New Zealand. This framework can account, in a simple way, for the mobile–immobile partitioning of fast- and slow-flow pathways of leaching through soil. The SPASMO simulations consider actively growing grapevines and a grassed interrow. The calculations are run on daily time-steps using a 27-year sequence of weather data (1972–99) recorded from the local airport at Napier. The soil's physical and hydraulic properties are taken from the New Zealand Soils Database. Three herbicides, simazine, linuron, and amitrole, are 'applied' at the label rate on 15 September each year. This modeling of leaching predicts if and how often the soil-solution concentration at 3 m might exceed the maximum allowable value (MAV) of 2 ppb of active ingredient.

Simazine is quite mobile because of its low exchange with the soil's organic carbon (eqn 6). Also it is persistent, with a long degradation half-life of $t_{1/2} = 50$ days. Linuron is less mobile because of greater exchange with the soil's organic matter, and it is less persistent, with $t_{1/2} = 33$ days. Amitrole is also less mobile because of low exchange, but it is much less persistent, with $t_{1/2} = 3$ days. Therefore, concerning the likelihood of leaching to groundwater, the risks are expected to be ranked simazine > linuron > amitrole. This expectation is borne out in the simulated leaching results, presented for a depth z of 3 m (Figure 6). The average simazine concentration in the soil solution leaching through Twyford silt loam, over a 30-year period, is predicted to be just 0.03 μg l^{-1}, although 10% of the time it will exceed 0.12 μg l^{-1}, or 6% of MAV. Simazine should be used with great caution, for it poses a greater leaching risk than either of the other two pesticides.

The risk of pesticide leaching also depends on the depth to groundwater. Figure 7 shows the model results for linuron, as a function of depth. Here the data are presented as a cumulative probability function of exceedence. During the time taken to leach deeper into the soil profile, there is degradation of the pesticide. For linuron, this means that there is predicted to be, on average (50% exceedence), about an order-of-magnitude drop in the concentration of active ingredient with every meter of soil depth. Even in the worst case, $z = 1$ m, the concentration of linuron will still be approximately an order of magnitude less than the MAV. None the less, this analysis shows that shallow, unconfined groundwaters are at the greatest risk of contamination from the leaching of surface-applied chemicals.

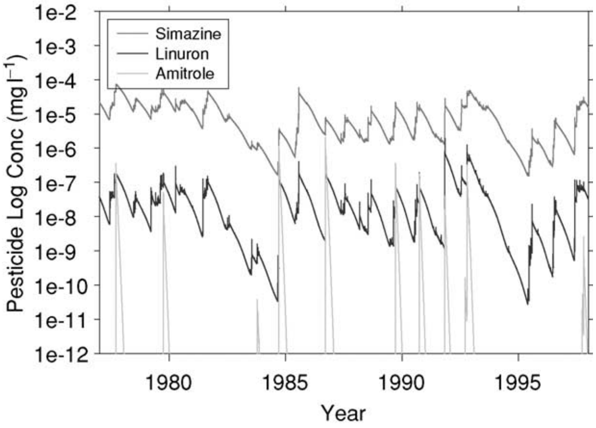


Figure 6 Simulated herbicide concentrations in the soil solution at 3 m depth under grapevines growing on a Twyford silt loam. Each year, over 30 years, the herbicides were considered to be applied at the label rate, to 100% of the soil surface, in mid-September.

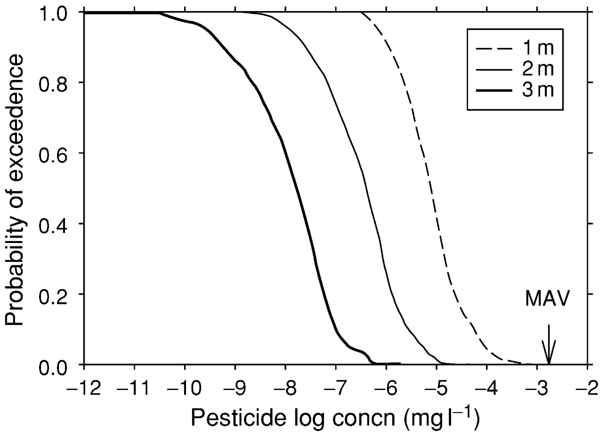


Figure 7 The predicted probability distribution for the concentration of linuron in leachate at three depths under grapevines growing on a Twyford silt loam. Linuron was considered to be applied once each year, in mid-September, at the label rate of 1.0 kg ha^{-1} of active ingredient. MAV, maximum allowable value.

Summary

The leaching of chemicals through soil by percolating pore-water is the result of a complex of interconnected biophysical and biochemical processes. Firstly, biophysical mechanisms determine the quantity of water that quits the root zone, destined for either groundwater or surface-receiving waters. The prevailing weather, the nature of the vegetative cover, and the properties of the soil conspire to determine the amount of leachate. Secondly, the biochemical processes of exchange, production, and degradation control the concentration of chemical in the leaching solution. Risk assessment of leaching needs to consider both the chemical concentration of the leachate solution and the total loading of chemical, which is the product of leachate quantity and leachate quality.

See also: **Diffusion**; **Evapotranspiration**; **Infiltration**; **Irrigation:** Environmental Effects; **Macropores and Macropore Flow, Kinematic Wave Approach**; **Pesticides**; **Plant–Soil–Water Relations**; **Pollution:** Groundwater; **Rhizosphere**; **Root Architecture and Growth**; **Soil–Plant–Atmosphere Continuum**; **Solute Transport**; **Sustainable Soil and Land Management**;

Vadose Zone: Hydrologic Processes; **Water Potential**; **Water-Use Efficiency**

Further Reading

Clothier BE and Green SR (1997) Roots: the big movers of water and chemical in soil. *Soil Science* 162: 534–543.

Clothier BE, Kirkham MB, and MacLean JE (1992) *In situ* measurement of the effective transport volume for solute moving through soil. *Soil Science Society of America Journal* 56: 733–736.

Clothier BE, Green SR, Roygard JFK, Vogeler I, and Duwig C (2000) *Modeling Rootzone Processes: a Tool for Environmental Risk Assessment*. Proceedings of the *Water 2000 Conference*, NZ Water and Wastes Association, Auckland, 19–23 March.

Feddes RA, Kowalik PJ, Malinka KK, and Zradny H (1976) Simulation of field water uptake by plants using a soil water dependent root extraction function. *Journal of Hydrology* 31: 13–26.

Gardner WR (1960) Dynamic aspects of water availability to plants. *Soil Science* 89: 63–73.

Green SR, Clothier BE, Mills TM, and Millar A (1999) Risk assessment of the irrigation requirements of field crops in a maritime climate. *Journal of Crop Production* 2: 353–377.

Green SR, Clothier BE, Caspari HW, and Neal SM (2002) Rootzone processes, tree water use and the equitable allocation of water to olives. In: Raats PAC, Smiles DE, and Warrick AW (eds) *Heat and Mass Transfer in the Natural Environment*. American Geophysical Union Monograph. Washington, DC: American Geophysical Union Books.

Jarvis NJ (1994) *The MACRO Model (version 3.1) – Technical Description and Sample Simulations*. Reports & Dissertations 19. Uppsala: Swedish University of Agricultural Science.

Jaynes DB and Horton R (1997) Field parameterization of the mobile/immobile domain model. In: Selim M and Ma L (eds) *Physical Nonequilibrium in Soils: Modeling and Application*. Ann Arbor, MI: Ann Arbor Press.

Jury WA and Roth K (1990) *Transfer functions and solute movement through soil: theory and applications*. Basel: Birkhauser.

Scotter DR and Kanchanasut P (1981) Anion movement in a soil under pasture. *Australian Journal of Soil Research* 19: 299–307.

Van Genuchten MT and Wierenga PJ (1976) Mass transfer studies in sorbing porous media. I. Analytical solutions. *Soil Science Society of America Journal* 40: 473–480.

LIEBIG, JUSTUS VON

R R van der Ploeg, University of Hannover, Hannover, Germany
W Böhm, Georg-August-University, Göttingen, Germany
M B Kirkham, Kansas State University, Manhattan, KS, USA

Introduction

Justus von Liebig (1803–73) was one of the most productive, significant, and versatile German chemists of all times. Because his contributions to the various fields of organic, analytical, agricultural, physiological, industrial, pharmaceutical, and educational chemistry are so basic and numerous, his influence is recognized even today. He is particularly celebrated for his chemicophysiological work on the metabolism of organisms and on the chemistry of agricultural soils. He is often considered as one of the founders, if not as the founder, of agricultural chemistry and as the father of modern agronomy. He is one of the most decorated scientists of all times and continues to be commemorated throughout the world. In 1845, for example, at the age of only 42, he was ennobled (which explains the 'von' in his name after 1845). The front of the beautiful Main Library of Iowa State University at Ames, Iowa, USA, shows the name of Liebig amidst those of other all-time celebrities of science and literature such as Newton, Pasteur, Faraday, Darwin, Shakespeare, and Emerson. The International Union of Soil Sciences (IUSS) in 2002 decided at its Congress in Bangkok to call one of its major awards after Liebig. In Germany, in almost every sizeable town, there is a street that is named after Liebig. The Justus-Liebig-Universität in Giessen, where he was a professor from 1824 to 1852, has borne his name since 1946.

Liebig's engagement as an agricultural chemist started in 1840, when he published his classic work *Organic Chemistry in its Application to Agriculture and Physiology*. Based on an analysis of the literature, Liebig rejected in this publication the so-called humus theory, with which hitherto the nutrition of plants was explained. Instead, he presented the theory on the mineral nutrition of plants, which implies that plants need, besides CO_2 from the air, mineral salts from the soil to develop. At the same time, Liebig recommended the use of mineral fertilizers in crop production.

In recent years, however, it has been shown that the agronomist and chemist Carl Sprengel (1787–1859) was the first to disprove the humus theory (1826), to formulate the theory of the mineral nutrition of plants (1828), and to recommend the use of mineral fertilizers (1828). Justus von Liebig was familiar with Sprengel's work, but did not cite his countryman's pioneering achievements and thus has to be blamed for plagiarism. Nevertheless, as an indefatigable leader of agronomy in the struggle for the acceptance of the theory of mineral nutrition of plants and the use of mineral fertilizers in crop production, as a reformer of agricultural education, and as a founder of the experimental and research station system, Liebig deserves to be commemorated as one of the most outstanding scholars of agriculture.

Life History

Justus Freiherr (Baron) von Liebig was born on 12 May 1803 in Darmstadt (Hesse) as Johann Justus Liebig, and he died on 18 April 1873 in Munich. He was the second son of Johann Georg Liebich (Liebig) and his wife Marie Caroline Fuchs. His father owned a drugstore in Darmstadt and sold paints and varnishes that he produced himself. The young Liebig gained his first chemical experience in his father's laboratory. At the age of 14, Justus Liebig dropped out of high school and worked for almost 2 years with his father. During this time, he started an apprenticeship in a pharmacy, which he did not complete. Although he had not finished high school, he registered at the age of only 16 as a university student, first at Bonn (Figure 1) and, from 1821, at Erlangen. There he worked, as he had in Bonn, under the supervision of the chemist Karl Wilhelm Gottlieb Kastner. In 1822 Liebig left for Paris to become a coworker of the famous chemist L.J. Gay-Lussac. In 1823 he earned, in his absence (because of political turmoils in Germany), his doctorate from the University of Erlangen with a thesis, 'The relation between mineral chemistry and plant chemistry.' One year later, at the age of only 21, Liebig accepted a professorship at the University of Giessen (Figure 1), and in 1825 he became a full professor of chemistry at the same university. In 1826 he married Henriette Moldenhauer. The Liebigs had five children, two boys and three girls.

At Giessen, Liebig established a laboratory to teach chemistry through research and experimentation. Within a few years, this unique laboratory developed into the most significant teaching and research facility in Germany. Because of his exceptional talent as a teacher and his epoch-making research, Liebig soon became a world-famous chemist. He attracted students from all over the world to Giessen. Among his German students were such chemists as Hoffmann, Kekulé, Pettenkofer, Fresenius, and Erlenmeyer. One of his British students was Joseph Henry Gilbert,

Figure 1 Germany, showing present and previous borders.

later coworker of John Bennet Lawes at Rothamsted. Of Liebig's American students, E.N. Horsford, J.A. Porter, and S.W. Johnson deserve special mention. After completing their studies in Liebig's laboratory at Giessen, Eben N. Horsford accepted a professorship at the Lawrence Scientific School at Harvard, and John A. Porter became a Professor of Analytical and Agricultural Chemistry at Yale University. Samuel W. Johnson also became a Professor of Agricultural Chemistry at Yale and, in addition, established the Connecticut Agricultural Experiment Station, the first institution of this kind in the USA. Later Johnson's influence did much to secure the passage of the so-called Hatch Act in 1887, which enabled the establishment of at least one agricultural experiment station in every state of the USA.

After 28 years at the University of Giessen, in 1852 Liebig accepted an invitation from King Maximilian II of Bavaria to go to the University of Munich. Here a new research institute was established, completely in accordance with Liebig's ideas and aspirations. In contrast to Giessen, this research laboratory was exclusively for Liebig and his assistants, not students. Liebig's picture of 1853 is shown in Figure 2. In 1859 Liebig was elected as President of the Royal Academy of Sciences, in Munich, an honorary position he held until his death in 1873. One of his sons, Hermann von Liebig (1831–94), became a respected agricultural chemist himself. Hermann worked for many years at the Agricultural Experiment and Research Station in Munich and published significant work. The other son, Georg von Liebig (1827–1903), who was a medical clinician, published after his father's death hitherto unpublished material, such as speeches by his father and correspondence with renowned colleagues.

Figure 2 The chemist Justus von Liebig (1803–73), who entered the field of agronomy in 1840 with his classic *Agrikulturchemie*. Engraved by A. Weger (Leipzig) from a photograph taken in 1853.

The Work of Liebig

From 1822 on, Liebig published extensively, at first mainly on topics of inorganic chemistry, and, after 1830, increasingly on organic chemistry and pharmacy. By the end of the 1830s, Liebig was a distinguished and world-famous scientist, with nearly 300 much-regarded, high-impact publications in the leading journals of chemistry and pharmacy. It was at this point that Liebig entered the field of agriculture almost by accident. In 1837, during a visit to Britain, Liebig was asked by the President of the British Association for the Advancement of Science to give a report on the state of organic chemistry and organic analysis. Liebig wrote this report, which was published in French in three parts in 1840 in Paris. The introduction to this report appeared in the same year as a book in German under the title *Die organische Chemie in ihrer Anwendung auf Agricultur und Physiologie* (*Organic Chemistry in its Application to Agriculture and Physiology*).

In this book (with the short title *Agrikulturchemie*), Liebig reviewed and analyzed current knowledge of plant physiology and plant nutrition. He especially criticized the so-called humus theory, which stated that plants need humus to develop. In studying the literature, Liebig concluded that plants did not need humus to grow, but they required CO_2 from the air and, in particular, mineral salts from the soil. For that reason Liebig recommended that farmers needed to replace all the mineral nutrients that were removed from their fields at harvest, to maintain the soil's fertility.

Because of the boldness with which Liebig demanded a change in fertilization, his book received much attention, from both farmers and agronomists. In a short time, five further editions appeared. From the fifth edition on, the title was slightly modified (*Die Chemie in ihrer Anwendung auf Agricultur und Physiologie*). The doctrines posed by Liebig gained ardent consent as well as sharp rejection. His claim to abandon organic fertilizing was questioned; also, the mineral fertilizers that he recommended were looked at with skepticism. In particular, his 'patent' fertilizer did not show the predicted yield increases. It probably failed because Liebig used highly insoluble mineral compounds; he assumed that soluble compounds would be leached from the soil too easily.

Another matter of skepticism, which hindered a quick breakthrough of the theory of the mineral nutrition of plants, was the unsolved question of the need to fertilize field crops with nitrogen. In the fifth edition of his *Agrikulturchemie* of 1843, Liebig said that nitrogen fertilization was not necessary, because crops would receive ample ammonia from the atmosphere. However, this statement contradicted farm experience; most crops had increased yields when fertilized with nitrogen. The strongest opposition to Liebig's view came from England, where John Bennet Lawes and his coworker Joseph Henry Gilbert had been carrying out field experiments since 1843 at Rothamsted. Their main objective was to disprove Liebig's theory of nitrogen fertilization.

Because his patent fertilizer failed, for the next 10 years Liebig did not take part in public discussions concerning unsolved problems of agricultural chemistry. In 1855 he reentered the ongoing debate. In a basic contribution in the *Zeitschrift für deutsche Landwirthe* (*Journal for German Farmers*), Liebig sharply criticized the work of Lawes and Gilbert and simultaneously formulated, in 50 precise theses, his own doctrines of plant nutrition. This work was published, in somewhat modified form, as a book with the title *Die Grundsätze der Agrikultur-Chemie mit Rücksicht auf die in England angestellten Untersuchungen*. In addition, it was published in English, also in 1855, under the title *The Relations of Chemistry to Agriculture and the Agricultural Experiments of Mr J.B Lawes*. With this publication Liebig tried to deflect the remaining resistance against the theory of the mineral nutrition of plants. He also reluctantly slightly changed his views on the usefulness of nitrogen fertilization of field crops.

The second edition of this book (1855) also contains Liebig's first exact formulation of the so-called Law of the Minimum. It states that the yield of a crop in a field (of a given condition and composition) is directly related to a particular nutrient, which is indispensable for complete development of the plant and which is available in the soil (in proper form and quality) in the smallest quantity (in 'minimo'). For many years it was believed that the Law of the Minimum was first stated by Liebig. A study of the literature has shown, however, that Liebig's countryman Carl Sprengel had conceived this law as early as 1828. Today it is generally accepted that Liebig derived the law, as well as much of his theory on the mineral nutrition of plants, from the work of Sprengel, which was published in the 1820s and 1830s.

Liebig's *Agrikulturchemie* initiated a public dispute of many years, because of his thesis that field crops did not need to be fertilized with nitrogen. In his book of 1856 *Über Theorie und Praxis in der Landwirthschaft* (*About Theory and Practice in Agriculture*), published in Braunschweig, Liebig again sharply attacked his opponents (the 'nitrogenists'). Main adversaries in Germany were Carl Sprengel, Julius Stöckhardt, and Emil Wolff, and, in England, Lawes and Gilbert. Carl Sprengel discussed the use of mineral nitrogen as a fertilizer as early as 1828.

During this period of ample and vocal agricultural controversies, Liebig published in 1858 his paper *Über das Verhalten der Ackerkrume zu den in Wasser löslichen Nahrungsstoffen der Pflanzen* (*About the Behavior of the Mold with Respect to the Water-Soluble Nutrients of Plants*). In this publication Liebig discussed the (cation) exchange capacity of arable soils, as observed around 1850 by John Thomas Way. From Way's work Liebig concluded that effective mineral fertilizers should be partly water-soluble. Based on this new insight, he was now able to explain the failure of his earlier patent fertilizer, which was highly insoluble. In 1859, with this new knowledge, he tried, by means of another book called *Naturwissenschaftliche Briefe über die moderne Landwirthschaft* (*Letters of Natural Sciences About Modern Agriculture*), to convince land users to practice farming solely on the basis of information derived from natural sciences.

Motivated by accumulated knowledge, Liebig revised in 1862 his *Agrikulturchemie*. The book now consisted of two volumes with different subtitles: volume 1 was called *Der chemische Prozess der Ernährung der Vegetabilien*, and volume 2, *Die Naturgesetze des Feldbaus* (*The Chemical Process of the Nutrition of Plants,* and *The Natural Laws of Crop Production*). By comparing previous editions, it is apparent that Liebig stuck to his earlier agricultural chemical principles. However, he did add new pieces of knowledge garnered since the appearance of his preceding editions. He also accepted some farming practices, gained by long-term experience rather than by scientific reasoning or modern agricultural experimentation. He made a small concession to his opponents by admitting that, under certain conditions, nitrogen fertilization might be necessary.

Liebig's colleagues generally gave positive evaluations of the seventh edition of his *Agrikulturchemie*. In a review for the *Agronomische Zeitung* of 1865, Wilhelm Hamm wrote, for example, that "with this book we have received an agricultural codex for all times." This appraisal seemed to be correct, because the seventh edition indeed listed almost all the facts of modern plant nutrition and fertilizer science. The seventh edition of *Agrikulturchemie* was not the last one. In 1865 the eighth edition appeared, and, in 1876, three years after Liebig's death, Philipp Zöller published a ninth edition. *Agrikulturchemie*, a classic work in the history of science, was translated into many foreign languages, especially in the early years after its first appearance. By 1848, the book had passed through more than 20 editions and reprints: six in Germany; five in England; three in the USA; two in France; two in Italy; and one each in Denmark, the Netherlands, Sweden, Poland, and Russia.

As an agricultural chemist, Liebig also dealt with questions regarding the nutrition of animals. His most significant publication in this area is the book *Die organische Chemie in ihrer Anwendung auf Physiologie und Pathologie* (*Organic Chemistry in Its Application to Physiology and Pathology*), which was published in Braunschweig in 1842, with a second edition in 1843 and a third one in 1847.

Liebig also considered problems of agricultural chemistry in his *Chemische Briefe* (*Chemical Letters*), published in Heidelberg in 1844. This book had a number of editions; the sixth edition was published in both Heidelberg and Leipzig. Also Liebig had contributions dealing with agricultural chemistry in the journal *Annalen der Chemie und Pharmacie*, of which he was an editor. Additionally, Liebig published frequently in agricultural journals about fertilizers and plant nutrition. Of much interest for the history of agronomic sciences is the correspondence between Justus von Liebig and Theodor Reuning, a highly influential agricultural administrator in Saxony (Germany). This correspondence was published by Reinhold Echtermeyer and Georg von Liebig in 1884 in Dresden and was titled *Briefwechsel zwischen Justus von Liebig und Theodor Reuning über landwirthschaftliche Fragen aus den Jahren 1854–1873*.

Liebig's merits as a reformer of agricultural education are important. From 1860 discussions had been taking place in Germany as to how to improve agricultural education. Up to then, advanced education in agriculture was almost exclusively offered by isolated agricultural academies. In two public addresses during meetings of the Royal Academy of Sciences in 1861 in Munich, Liebig called for the closing of existing agricultural academies and, instead, for the establishment of educational programs in agriculture at universities. He felt that only at universities was it possible to teach adequately the basic sciences that agriculture needed. His addresses were published by his son Georg von Liebig and M. Carrière in 1874, in Leipzig and Heidelberg, as *Justus von Liebig: Reden und Abhandlungen*. Liebig's efforts to implement the study of agriculture at universities resulted in the creation of agricultural departments at a number of universities and colleges.

Last, but not least, Liebig's influence on the founding of agricultural extension programs in Germany and abroad needs to be mentioned. With much success, he convinced governmental authorities at home and in other European countries of the necessity of providing funds and facilities for the establishment of agricultural experiment stations. The first state agricultural experiment station in Germany was established at Möckern (now a suburb of Leipzig) in Saxony in 1851–52. Twenty-five years later there were 74 such stations in Germany, 16 in Austria, and 10 in Italy. At each station a chemist with one or more assistants was engaged in the analysis of food, feedstuff, fertilizers, and other agricultural products, and in conducting experiments on the fertilization of crops and the feeding of animals. Liebig and his school based agriculture so overwhelmingly on chemistry that, well after the beginning of the twentieth century, agricultural chemistry was considered synonymous with agricultural science.

Liebig and Sprengel

A discussion about the significance of Justus von Liebig as an agricultural chemist is not complete without a reference to Liebig's countryman Carl Sprengel. Philipp Carl Sprengel was born in 1787 in Schillerslage (today a suburb of Burgdorf) near Hannover in Northern Germany (Figure 1), and he died in 1859 in Regenwalde, now Resko (Poland). As a student of Albrecht Daniel Thaer (the founder of modern agriculture) and as a farm manager for many years, he was one of the most experienced and knowledgeable agronomists in Germany, when he registered in 1821, at the age of 34, as a student of natural sciences at the University of Göttingen. After earning his doctorate in 1823 (Justus von Liebig was awarded his doctorate in the same year), Sprengel worked eight more years at the University of Göttingen, before he transferred to Braunschweig in 1831 to become a college professor. From there he moved to Regenwalde (Figure 1), where he founded his private academy of agriculture. Sprengel's undated picture is shown in Figure 3.

Sprengel carried out pioneering research in Göttingen between 1823 and 1831. His first publications dealt with the humus theory and the role of soil organic matter (humus) as the only source of plant nutrients. He analyzed the water-soluble constituents in the humus extract and found a variety of salts such as alkali nitrates, sulfates, chlorides, and phosphates. Because of the salts that he measured in the ash of plants, Sprengel concluded that these soluble salts in the humus extract were the real plant nutrients. Sprengel published his results (with which he essentially disproved the humus theory) in 1826 in the *Archiv für die gesammte Naturlehre*. The editor of this journal was K.W.G. Kastner, under whose

Figure 3 The agronomist and chemist Carl Sprengel (1787–1859), who disproved, in 1826, the humus theory and who published in 1828 the theory of the mineral nutrition of plants, as well as the Law of the Minimum. (Undated picture from the Göttingen State and University Library, Göttingen, Germany.)

guidance Justus von Liebig had earned his doctorate in 1823. Initially, the work of Liebig was also published in this journal.

Sprengel's most important publication was probably Von den Substanzen der Ackerkrume und des Untergrundes (About the substances in the mold and the subsoil), which appeared in the *Journal für Technische und Oekonomische Chemie in 1828*. In this lengthy article, Sprengel explained in detail and precisely his theory on the mineral nutrients of plants. In this article he also stated the Law of the Minimum. In later publications, for example in Sprengel's books on agricultural chemistry of 1831 and 1832, the theory of the mineral nutrition of plants and the Law of the Minimum are presented again. Sprengel was a most productive scholar in the first half of the nineteenth century. Besides a large number of journal articles, he published fundamental books on agricultural chemistry, soil science, and crop production. As early as 1828, he had firm ideas about fertilizing field crops with mineral nitrogen. In his article of 1828, he discussed in detail the usefulness of $(NH_4)NO_3$, $(NH_4)_2SO_4$, $(NH_4)Cl$, and $(NH_4)_2CO_3$, as well as KNO_3, $NaNO_3$, and $Ca(NO_3)_2$ as mineral fertilizers.

Liebig was very familiar with the work of Sprengel; but he refused to acknowledge Sprengel as the scholar who had first published the theory of the mineral nutrition of plants and the Law of the Minimum, and who had first recommended the use of mineral fertilizers in crop production. There was a passionate dispute between Sprengel and Liebig about parts of Liebig's *Agrikulturchemie* of 1840, which were taken from Sprengel's work without proper acknowledgement. Liebig, however, who is also remembered as an ambitious, highly argumentative, and intolerant man, ignored Sprengel's justified complaints and continued to pass off the theory of the mineral nutrition of plants as his own. Because of his inappropriate behavior toward Sprengel, and also because of innumerable and harsh disputes with other colleagues, Liebig was not particularly liked by his peers. Some of his colleagues, like the Dutch chemist and physiologist G.J. Mulder from the University of Utrecht, showed their antagonism publicly. (In a widely circulated brochure, Mulder wrote, in 1846: "Freedom of scientific opinion has never been understood by Liebig. For years past a tribunal has been established in Giessen, before which Liebig is at the same time accuser, witness, public prosecutor, advocate, and judge. Before this tribunal a case is rapidly terminated, but grace or justice can never, never be obtained there. From this tribunal, even the purest innocence is never dismissed without being whipped and branded; and for this purpose Liebig holds also the office of executioner.")

Around the middle of the twentieth century, Günter Wendt, from the University of Göttingen, made a thorough study of Carl Sprengel and his pioneering work in soil chemistry and agronomy. Wendt's work about Sprengel and Liebig caused a good deal of confusion in Germany. In 1955, the influential Association of German Agricultural Experimental and Research Stations (VDLUFA) decided to take proper action. It created the Sprengel–Liebig Medal in gold (Figure 4), with which persons are regularly honored for outstanding achievements in, or services to, agriculture. By doing so, a dispute about priorities and impacts was avoided, and both pioneering scholars, Sprengel and Liebig, were equally recognized and commemorated.

Summary

Liebig, who spent more than 30 years in intensive engagement in agricultural chemistry, has influenced, since 1840, the course of agronomic developments in Germany and many other countries, particularly in the USA. Because he expounded on the doctrine of the mineral nutrition of plants, it became generally accepted. He promoted the production and use of mineral fertilizers in crop production. The establishment of numerous agricultural experiment stations during the second half of the nineteenth century in Germany and abroad was, in large part, a result of his efforts. More than anybody else, Liebig called for the use of scientific principles in agriculture. He initiated

Figure 4 The gold Sprengel–Liebig medal, established in 1955 by the Association of German Agricultural Experimental and Research Stations (VDLUFA), to commemorate and honor both pioneering agronomists.

a reform in the educational system of agronomists and made agriculture a university discipline. Occasionally, as in the case of nitrogen fertilization, he advocated incorrect views, but this does not affect his outstanding position in the history of the agricultural sciences. However, he should be commemorated in agriculture as a most prominent leader and reformer, and as an untiring propagandist of using mineral fertilizers in crop production, rather than as a discoverer of new knowledge. To Carl Sprengel, on the other hand, belongs the honor of having disproved the humus theory and of having conceived both the theory of mineral nutrition of plants and the Law of the Minimum.

Further Reading

Böhm W (1997) *Biographisches Handbuch zur Geschichte des Pflanzenbaus (Biographical Handbook on the History of Agronomy)*. Munich, Germany: Saur-Verlag.

Browne CA (1944) A source book of agricultural chemistry. In: Verdoorn F (ed.) *Chronica Botanica*, vol. 8, no. 1. Waltham, MA: Chronica Botanica.

Dyke GV (1993) *John Lawes of Rothamsted – Pioneer of Science, Farming and Industry*. Harpenden, UK: Hoos Press.

Liebig J (1840) *Die organische Chemie in ihrer Anwendung auf Agricultur und Physiologie (Organic Chemistry in its Applications to Agriculture and Physiology)*. Braunschweig, Germany: Friedrich Vieweg.

Liebig J von (1855) *Die Grundsätze der Agricultur-Chemie mit Rücksicht auf die in England angestellten Untersuchungen (The Relations of Chemistry to Agriculture and the Agricultural Experiments of Mr JB Lawes)*, 1st and 2nd edn. Braunschweig, Germany: Friedrich Vieweg.

Moulton FR (1942) *Liebig and after Liebig*. Publication No. 16. Washington, DC: American Association for the Advancement of Science.

Paoloni C (1968) *Justus von Liebig: Eine Bibliographie sämtlicher Veröffentlichungen*. Heidelberg, Germany: Carl Winter University Press.

Rossiter MW (1975) *The emergence of agricultural science: Justus von Liebig and the Americans, 1840–1880*. New Haven, CN: Yale University Press.

Sprengel C (1826) Ueber Pflanzenhumus, Humussäure und humussaure Salze (About plant humus, humic acids and salts of humic acid). *Archiv für die Gesammte Naturlehre* 8: 145–220.

Sprengel C (1828) Von den Substanzen der Ackerkrume und des Untergrundes (About the substances in the mold and the subsoil). *Journal für Technische und Ökonomische Chemie* 2: 423–474; 3: 42–99, 313–352, 397–421.

Sprengel C (1831) *Chemie für Landwirthe, Forstmänner und Cameralisten (Chemistry for Agronomists, Foresters, and Agricultural Economists)*, vol. 1. Göttingen, Germany: Vandenhoeck und Ruprecht-Verlag.

Sprengel C (1832) *Chemie für Landwirthe, Forstmänner und Cameralisten (Chemistry for Agronomists, Foresters, and Agricultural Economists)*, vol. 2. Göttingen, Germany: Vandenhoeck und Ruprecht-Verlag.

Thaer A (1809–1812) *Grundsätze der rationellen Landwirthschaft (The Principles of Agriculture)*, vols 1–4. Berlin, Germany: G Reimer. (Translated by Shaw W and Johnson CW.) New York: Greeley and McElrath.

van der Ploeg RR, Böhm W, and Kirkham MB (1999) On the origin of the theory of mineral nutrition of plants and the law of the minimum. *Soil Science Society of America Journal* 63(5): 1055–1062.

Wendt G (1950) *Carl Sprengel und die von ihm geschaffene Mineraltheorie als Fundament der neuen Pflanzenernährungslehre (Carl Sprengel and his Mineral Theory as Foundation of the Modern Science of Plant Nutrition)*. Wolfenbüttel, Germany: Ernst Fischer-Verlag.

LIMING

E J Kamprath and T J Smyth, North Carolina State University, Raleigh, NC, USA

Introduction

Approximately 40% of the arable land in the world is naturally acid. The naturally acid soils have developed where rainfall exceeds evapotranspiration. Acid soils have also developed with intensive agriculture. The acidity of soils can be divided into that associated with the soil solution, exchangeable, and nonexchangeable forms. The poor growth of plants on acid soils is due to toxicities of Al, Mn, and H, and deficiencies of Ca, Mg, and Mo. Liming of acid soils increases plant growth due to neutralization of toxic elements, supplying Ca and/or Mg and increasing availability of Mo and P. Determination of the lime requirement for a soil can be based on the amount needed to adjust pH to a desired level, neutralization of exchangeable Al, or the amount estimated by the buffer pH method. Selection of a liming material has to take into account the neutralization value of the lime, fineness of the material, and the need for Mg.

Development of Acid Soils

In nature acid soils are found where rainfall exceeds evapotranspiration. Several processes contribute to soil acidification under these conditions. The excess water moving through the soil carries Ca^{2+} and Mg^{2+} which balance the negative charge of soluble anions like NO_3^- and Cl^-. Leaching of Ca and Mg in conjunction with soluble anions decreases the percentage of exchange sites occupied by basic cations. The sites vacated by Ca and Mg are initially replaced with H^+ supplied by decomposition of organic matter and residues. The H^+-clays then decompose to form Al-clays.

With intensive agriculture H^+ is also produced by the conversion of ammonium to nitrate:

$$NH_4^+ + 2O_2 \rightarrow NO_3^- + H_2O + 2H^+ \qquad [1]$$

The application of 100 kg N ha^{-1} in the ammonium form can require a maximum of 360 kg $CaCO_3$ ha^{-1} to neutralize the acidity produced by the conversion of NH_4^+ to NO_3^-. Plants also release H^+ from roots when their uptake of basic cations exceeds that of anions. Since legumes fix atmospheric N and have relatively high concentrations of Ca, cation uptake exceeds anion uptake. To maintain internal plant electroneutrality legume roots excrete H^+ which acidifies the rhizosphere. Neutralization of acidity produced in one growing season of N-fixing legumes can correspond to 30–400 kg $CaCO_3$ ha^{-1}.

Another source of H^+ is acid precipitation produced upon burning fossil fuels containing sulfur compounds which give rise to SO_2. The SO_2 reacts with water in the atmosphere to produce sulfuric acid.

Nature of Soil Acidity

Soil Solution Acidity

The solutions of acid soils contain two ions that are detrimental to plant growth, hydrogen and aluminum. The concentration of Al^{3+} in the soil solution increases with the proportion of cation exchange sites which are occupied with Al (i.e., the percentage Al saturation of the cation exchange capacity (CEC)). There is a marked increase in soil solution Al^{3+} concentration when the saturation is greater than 50%. The soil solution H^+ concentration is controlled by the hydrolysis of Al^{3+}, as shown by eqn [2].

$$Al^{3+} + H_2O \rightleftarrows AlOH^{2+} + H^+ \qquad [2]$$

Between pH values of 5.6 and 5.8 the soil solution Al^{3+} concentration is nil in mineral soils since the Al saturation is essentially zero. Organic soils with appreciable mineral matter have very low concentrations of soil solution Al^{3+} at pH values of 4.8–5.0, because Al is strongly bound to the carboxyl groups. High concentrations of soil solution Al^{3+} only occur in organic soils at pH values of 4.5 or lower.

Exchangeable Acidity

Early concepts were that exchangeable H^+ was the source of soil acidity. Titration studies showed that acid clays behaved as weakly ionized acids, which led to the conclusion that acid soils contained exchangeable H^+. Soils were assumed to be 100% base-saturated at pH values of 7 and the difference between the CEC at pH 7 and the sum of basic cations was the amount of exchangeable H^+. Later studies revealed that the weakly ionized acid characteristic of clays was due to exchangeable Al and the hydrolysis of Al^{3+} was the source of soil solution H^+. When saturated with H^+ the clays decomposed, releasing Al, which occupied the exchange sites.

Exchangeable acidity is defined as that which is extracted with a neutral unbuffered salt solution such as KCl. Neutral unbuffered salt solutions must be used to avoid changing the soil pH and altering the

Table 1 The exchangeable and nonexchangeable acidity of acid soils

Soil	pH	Organic matter (g kg^{-1})	Exchange Al (cmol kg^{-1})	Al saturation (%)	Acidity: Titratable[a] (cmol kg^{-1})	Acidity: Nonexchangeable cmol kg^{-1}
Typic Kandiudult	4.5	20	0.91	82	1.70	0.57
Typic Paleaquult	4.7	27	1.04	78	1.72	0.68
Aeric Paleaquult	4.5	40	2.33	73	3.77	1.44
Typic Haplustox	4.5	–	1.55	68	3.85	2.70

[a] Acidity titrated to pH 6.

Data from Evans CE and Kamprath EJ (1970) Lime response as related to percent Al saturation, solution Al, and organic matter content. *Soil Science Society of America Proceedings* 34: 895–896. Kamprath EJ (1970) Exchangeable aluminum as a criterion for liming leached mineral soils. *Soil Science Society of America Proceedings* 34: 252–254. Gonzalez-Erico E, Kamprath EJ, Naderman GC, and Soares WV (1979) Effect of depth of lime incorporation on the growth of corn on an Oxisol of Central Brazil. *Soil Science Society of America Journal* 43: 1155–1158.

exchangeable acidity. In mineral soils the exchangeable acidity is Al^{3+}. At pH values of 5 and lower, over half of the active exchange sites are countered by Al^{3+} (Table 1). As previously discussed, exchangeable Al is essentially zero in mineral soils at pH 5.6 and the active cation exchange sites are 100% base-saturated.

Nonexchangeable Acidity

Soils also contain H which is not extractable with neutral unbuffered salt solutions, but is titratable. The nonexchangeable H is associated with carboxyl groups of organic matter, hydroxy Al, and hydrated oxides of Fe and Al. The nonexchangeable acidity of the three Ultisols in Table 1 was primarily due to organic matter. However, with the Oxisol, the nonexchangeable acidity was associated with the hydrated oxides of Fe and Al. Once the exchangeable Al is neutralized the acid buffering capacity of the soil is a function of the nonexchangeable acidity. Neutralization of the nonexchangeable acidity has little effect *per se* on plant growth, but is the source of the pH-dependent CEC. With increasing neutralization of the nonexchangeable acidity, there is an increase in soil pH, CEC, and the retention of the basic cations Ca, Mg, and K.

Acid Soil Constraints to Plant Growth

Hydrogen Toxicity

Poor plant growth in acid soils is usually associated with low soil pH, but the direct effects of H^+ are often confounded with changes in solubility of various elements affecting plant growth. Experiments in hydroponics, therefore, are often used to assess the direct effects of H^+. Roots growing in solutions at pH 4.6 or lower have visual symptoms of injury, namely stunted growth, brownish color, and little lateral root development. Increasing solution Ca concentration alleviates root injury due to H^+ (Table 2). At low pH and Ca concentrations root membranes are damaged, leading to loss of organic substrates and absorbed cations. Under soil conditions with pH below 4.6, root injury due to H^+ can be expected along with reduced plant uptake and possible deficiencies of Ca and Mg.

Table 2 Soybean tap and lateral root length after 12 days of exposure to solutions with different pH and Ca concentrations

Solution: pH	Solution: Ca (mmol l^{-1})	Root length: Tap (cm per plant)	Root length: Lateral (cm per plant)
4.0	0.2	1.1	0.0
	2.0	3.2	1.0
4.6	0.2	10.2	3.4
	2.0	22.2	36.3
5.5	0.2	27.8	40.1
	2.0	44.6	85.6

Data from Sanzonowicz C, Smyth TJ, and Israel DW (1998) Calcium alleviation of hydrogen and aluminum inhibition of soybean root extension from limed soil into acid subsurface solutions. *Journal of Plant Nutrition* 21: 785–804.

Aluminum Toxicity

Aluminum toxicity is a major cause of poor plant growth in acid soils. Current evidence indicates that Al interferes with various root growth processes: disruption of regulatory signals in root cap cells and interference with cell division in root apices, enzyme activities, DNA replication, and P availability at membranes. The immediate and visible evidence of Al toxicity is a reduction in root length which, in turn, limits plant access to soil water and nutrients. Consequently, plants growing in soils with toxic levels of Al are underdeveloped, sensitive to drought stress, and often present symptoms of multiple nutrient deficiencies.

Concentrations of soil solution Al that reduce crop growth are in the micromolar range. As illustrated in

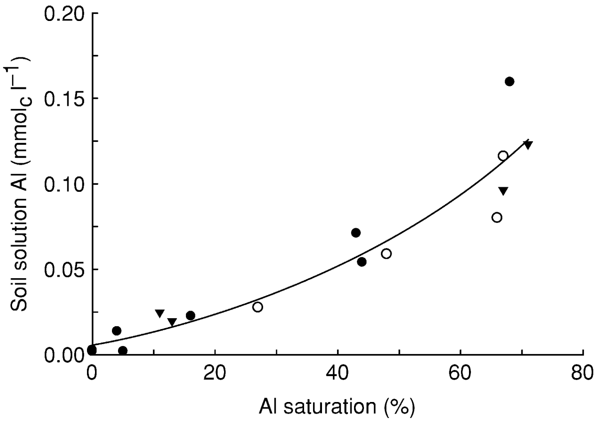


Figure 1 Relations between Al saturation of cation exchange sites and concentration of Al in the soil solution of Oxisols and Ultisols. Filled circles, Haplustox; open circles, Umbraquult; inverted triangles, Paleudult. Data source: Gonzalez-Erico E, Kamprath EJ, Naderman GC, and Soares WV (1979) Effect of depth of lime incorporation on the growth of corn on an Oxisol of Central Brazil. *Soil Science Society of America Journal* 43: 1155–1158; and Jallah JK (1994) *Assessment of Some Chemical Constraints to Root Growth in Four Acid North Carolina Soils.* PhD thesis. Raleigh, NC: North Carolina State University.

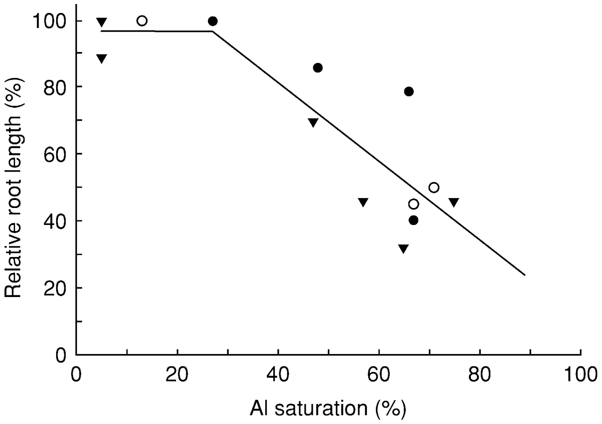


Figure 2 Effect of Al saturation on relative root length of field-grown corn in Oxisols and Ultisols. Filled circles, Umbraquult; open circles, Paleudult; inverted triangles, Haplustox. Data source: Gonzalez-Erico E, Kamprath EJ, Naderman GC, and Soares WV (1979) Effect of depth of lime incorporation on the growth of corn on an Oxisol of Central Brazil. *Soil Science Society of America Journal* 43: 1155–1158; and Jallah JK (1994) *Assessment of Some Chemical Constraints to Root Growth in Four Acid North Carolina Soils*. PhD thesis. Raleigh, NC: North Carolina State University.

Figure 1, the concentration of Al in soil solutions is related to the percentage of Al saturation of the cation exchange sites, with a marked increase in the Al concentration when the Al saturation exceeds 50%. Most field assessments of crop response to toxic levels of soil solution Al have been based on percentage of Al saturation as a proxy variable, due to its greater ease of measurement. Root length of corn, for example, is markedly reduced at percentage of Al saturation levels corresponding to high concentrations of Al in the soil solution (Figure 2).

Tolerance to Al varies widely among species and even varieties of the same species (Table 3). Yields for most varieties of cotton, mung bean, pearl millet, and wheat are reduced by low levels of percentage of Al saturation, whereas cassava and cowpea tolerate high levels of soil Al. However, considerable differences in Al tolerance/susceptibility have been observed among varieties within a given species. For example, critical percentage of Al saturation levels among field trials with different varieties range from 0 to 40 with corn, 0 to 44 with soybean, and 0 to 60 with upland rice.

Table 3 Threshold values of percentage Al saturation of the cation exchange capacity above which crop yield for various species is normally reduced by Al in mineral soils

Crop	*Critical % Al saturation*
Cassava	75
Corn	30
Cotton	0
Cowpea	60
Mung bean	0
Peanut	40
Pearl millet	0
Phaseolus bean	15
Potato	20
Sorghum	15
Soybean	15
Upland rice	40
Wheat	0

Data from Osmond DL, Smyth TJ, Reid WS *et al.* (2002) *Nutrient Management Support System, NuMaSS* (version 2.0). Raleigh, NC: Soil Management Collaborative Research Support System, North Carolina State University.

Manganese Toxicity

Acid soils with high contents of Fe and Al oxides often contain large amounts of soluble Mn. Plant growth in such soils is reduced when large quantities of the soluble soil Mn accumulate in their tissues. Although Mn is an essential nutrient, excessive plant uptake is toxic to plants. Liming these soils to pH 5.5 or greater decreases the solubility of soil Mn, which reduces the uptake of excess Mn and increases plant growth. In many instances, improved plant growth upon liming soils with percentage of Al saturation below the critical level is associated with the alleviation of Mn toxicities.

Calcium and Mg Deficiencies

Acid soils with high percentage of Al saturation have very low amounts of Ca and Mg on both the exchange sites and in the soil solution. Soils with very

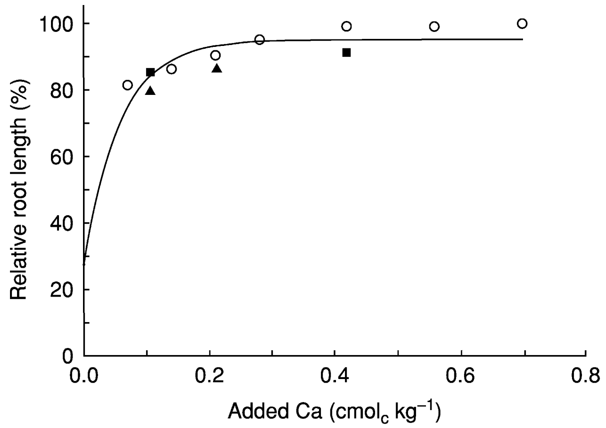


Figure 3 Root elongation of wheat as a function of Ca levels and sources in Ca-deficient Oxisols. Squares, lime; circles, $CaCl_2$; triangles, $Ca(H_2PO_4)_2$. Data source: Ritchey KD, Silva JE, and Sousa DMG (1983) Relação entre teor de calcio no solo e desenvolvimento de raizes avaliado por um metodo biologico. *Revista Brasileira Ciencia Solo* 7: 269–275.

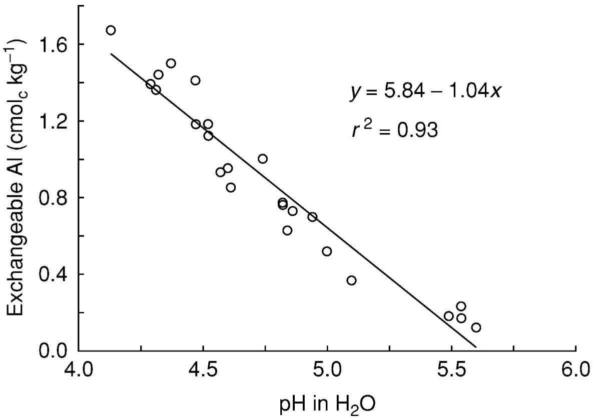


Figure 4 Relationship between soil pH and exchangeable Al for Coastal Plain soil. Data source: unpublished data of Soil Science Department North Carolina State University.

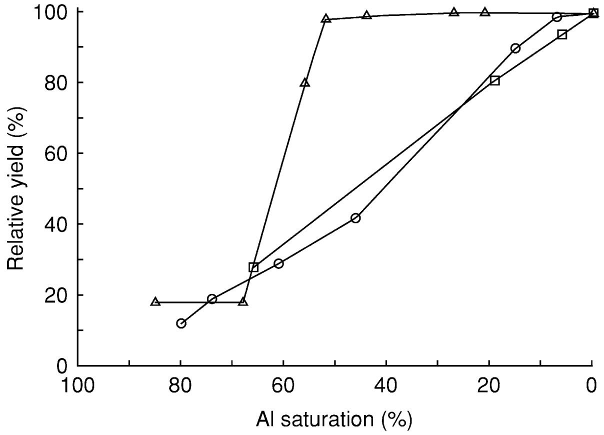


Figure 5 Relative yield of soybean (circles), corn (triangles), and cotton (squares) as related to Al saturation of the effective cation exchange capacity. Data source: E.J. Kamprath (unpublished data); and McCart GD and Kamprath EJ (1965) Supplying calcium and magnesium for cotton on sandy, low cation exchange soils. *Agronomy Journal* 57: 404–406.

low CEC have a limited capacity to retain Ca and Mg and supply of this nutrient may not be sufficient for normal plant growth. The immediate supply of Ca for continued growth of root tips is dependent on the soil solution concentration, because this nutrient is relatively immobile in plants. Root length of wheat seedlings, for example, was increased by over 50% when 0.1 cmol Ca kg^{-1} was added to an Oxisol which initially contained 0.02 cmol Ca kg^{-1} (Figure 3). Although it is often difficult to separate the extent to which a plant growth response to liming is due to correction of a Ca deficiency versus alleviation of an Al toxicity, the improved root growth of wheat was similar among various sources of Ca. Plant growth responses to additions of Mg are often associated with soil conditions where exchangeable Mg^{2+} and/or the proportion of cation exchange sites occupied by Mg^{2+} is very low.

Response to Liming

A major response of plants to liming is due to the neutralization of H, Al, and Mn, which if present in relatively high concentrations in the soil solution are toxic to plant growth. Toxic concentrations of H generally do not occur in soils except in certain organic soils when pH values are less than 4.5. In acid mineral soils with pH values less than 5.5 liming removes the negative effect of H on absorption of Ca, Mg, and K.

The response of plants to liming is mainly due to the neutralization of exchangeable Al and its replacement on the exchange sites with Ca and/or Mg. Liming of acid mineral soils to pH 5.6 reduces exchangeable Al to essentially zero (Figure 4). This essentially eliminates Al from the soil solution and removes it as a factor limiting plant growth. As previously discussed, the concentration of Al in the soil solution is a function of the Al saturation of the active cation exchange sites. Since the Al saturation of a soil is much easier to determine than soil solution Al, it is an excellent criterion for determining the need to lime a given soil. Since crops and genotypes vary in their tolerance to Al, knowledge about the Al saturation of a given soil provides information whether a particular crop or genotype will require liming. An example of this is the differential Al tolerance of corn as compared to soybean and cotton in a greenhouse study (Figure 5). Corn showed little effect to Al until the Al saturation was greater than 50%, where there is a marked increase in soil solution Al. Relative grain yields of corn on three soils decreased sharply when Al saturation was greater than 40% (Figure 6), a value similar to that found for dry-matter production in the greenhouse. Organic soils and soils high in organic matter have a lower critical pH than most mineral soils. Maximum plant growth on acid organic soils is generally obtained at pH 5. On these soils the exchangeable Al is held on carboxyl sites and at pH 5 the Al is held quite strongly so there is very little Al in the soil solution. Liming of these soils to pH 5 provides adequate Ca to overcome the H effects on plant growth.

Response to liming can also be obtained on acid soils with low Al saturation but that have high amounts of Mn. Soil with high contents of sesquioxides often contain appreciable amounts of Mn minerals and in these soils Mn toxicity is a problem when pH values drop to 5 and less. Liming to pH 5.6 and above eliminates Mn toxicity on these soils by converting the soluble Mn^{2+} to the relatively insoluble oxides, Mn_2O_2 and MnO_2.

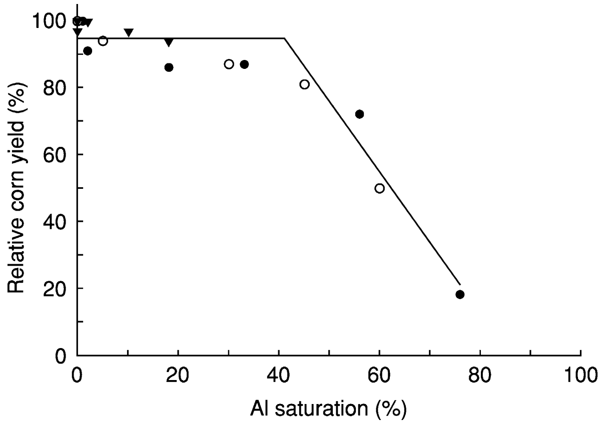


Figure 6 Relative grain yields of corn on three soils (filled circles, Goldsboro; open circles, Wharton; inverted triangles, Morrill) as related to exchangeable Al saturation of the effective cation exchange capacity. Data source: Alley MM (1981) Short-term chemical and crop yield responses to limestone applications. *Agronomy Journal* 73: 687–689; and Fox RH (1979) Soil pH, aluminum saturation and corn grain yield. *Soil Science* 127: 330–334.

Response to liming due to supplying of Ca *per se* is not often the case. In most instances Ca saturations of 20–25% will supply adequate Ca for plant growth except where toxicities to Al, H, or Mn limit plant growth. There is a good relationship between Ca or Ca + Mg saturation and plant growth, because these are the mirror image of Al saturation. On soils with very low CECs a minimum of 1 cmol Ca kg^{-1} soil is needed for optimum growth of most plants. When toxicities are not a factor the percentage of Ca saturation of the CEC is a good indicator of Ca availability. Calcium concentrations of a given plant species are similar when grown on soils with the same Ca saturation even though the soils contain a wide range of exchangeable Ca (Table 4). This is because the Ca concentration of the soil solution is determined by the percentage of Ca saturation rather than the amount of exchangeable Ca. Calcium is transported to the root by mass flow of water and thus the amount arriving at the root surface is a function of the Ca concentration. Magnesium deficiencies on sandy soils are generally associated with pH values of less than 5. The concentration of Al^{3+} and H^+ at pH 5 in mineral soils is sufficient to inhibit Mg uptake by plants. Liming to neutralize Al and H is necessary for optimum uptake of Mg where soil reserves are adequate. On acid sandy soils response to liming is also related to the supply of Mg with the application of dolomitic lime. Availability of soil Mg is related more to the Mg saturation of the CEC rather than the amount of exchangeable Mg. Various studies have shown that Mg saturation needs to be in the range of 5–10% for optimum plant growth.

Soybean and other legumes often show a response to increase in pH beyond that at which Al is neutralized. Raising the pH of mineral soils to 6 has

Table 4 Effect of Ca saturation of the effective cation exchange capacity (ECEC) on Ca availability

		Soil Ca			
Soil	*pH*	*ECEC saturation (%)*	*Exchangeable ($cmol_c$ kg^{-1})*	*Solution ($cmol_c$ l^{-1})*	*Soybean Ca (%)*
Norfolk	4.9	4	0.06	0.75	0.39
	5.8	72	1.29	4.00	0.86
Portsmouth	4.8	26	3.00	2.92	0.61
	5.3	77	8.31	5.00	0.99
Organic	4.2	31	7.08	3.75	0.47
	4.7	70	31.75	5.12	0.98

Data from Evans CE and Kamprath EJ (1970) Lime response as related to percent Al saturation, solution Al, and organic matter content. *Soil Science Society of America Proceedings* 34: 895–896. Evans CE (1968) *Ion Exchange Relationships of Aluminum and Calcium in Soils as Influenced by Organic Matter.* PhD thesis. Raleigh, NC: North Carolina State University.

increased yield because of increased solubility of soil Mo, which is required by bacteria for N_2 fixation. Increasing the OH^- concentration with liming brings about the replacement of MoO_4^{-2} absorbed by the Fe and Al hydrated oxides. Raising the pH to 6.0–6.2 makes sufficient Mo available for optimum N_2 fixation and optimum yields of legumes. Thus the higher critical pH for legumes as compared to nonlegumes on mineral soils is related to the Mo requirement for N_2 fixation and the effect of pH on Mo availability.

Availability of soil P in acid soils has been increased by liming. In acid soils Al at the root surface precipitates P and decreases the amount of P transported to plant tops. Removal of Al from the soil solution by liming prevents this from happening. Neutralization of exchangeable Al also results in more root growth and exploration of a greater soil volume. Phosphorus gets to roots by diffusion and when root growth in a given volume is increased, the uptake of P is increased.

The increased root growth in a given volume of soil when Al is neutralized by liming has a number of general beneficial effects on plant growth. Nutrient and soil water availability are increased. Microbial activity is enhanced; this affects N_2 fixation, decomposition of organic matter and plant residues, making available organically bound nutrients.

Cation Exchange Properties Affecting Liming of Acid Soils

The CEC of a soil is largely determined by the mineral composition of the clay-sized fraction. The source of the negative charge in clay minerals can be either permanent or variable. Permanent charge results from isomorphous substitution among cations with differing valences within the crystalline clay structure, and is a permanent characteristic of the mineral. Smectite, vermiculite, and illite are commonly occurring clay minerals with permanent charge. Variable charge results from the protonation and deprotonation of OH groups on the surfaces of Al and Fe oxides, crystalline or amorphous, and exposed edges of layer-silicate minerals. Kaolinite, hydroxy-interlayered vermiculite, goethite, gibbsite, quartz, and allophane are some of the clay minerals with variable charge. Carboxyl and OH groups in soil organic matter also have a variable-charge behavior, wherein the amount of negative charge increases with soil pH.

Most acid soils have a mixed composition of both permanent- and variable-charge clay minerals. However, the dominant type of clay mineral charge can be estimated through the sum of basic cations and exchangeable acidity (unbuffered salt extraction) relative to the amount of soil clay. Soils dominated by minerals with permanent charge have CEC values $>12\,\mathrm{cmol\,kg^{-1}}$ of clay, whereas soils dominated by variable-charge clays have CEC values $<12\,\mathrm{cmol\,kg^{-1}}$ of clay.

When lime is applied to soils with variable-charge clay minerals, part of the potential base reaction is consumed in the neutralization of H^+ released from OH groups. Consequently, the lime required to neutralize a targeted amount of exchangeable acidity is increased.

Lime Requirement Determinations

Soil pH

Soil pH was one of the first methods for determining the need for lime. At pH 7 soils were considered to be 100% base-saturated and had zero base saturation at pH 4. The relationship between soil pH and base saturation was considered to be linear between pH 4 and 7. Measurement of the soil pH provided an estimate of the base saturation. To determine the amount of lime to raise the soil to the desired soil pH requires information about the CEC of the soil which can be estimated based on soil texture and organic matter content. As an example, the amount of lime to bring the pH to 6 (90% base saturation) for a soil which had an initial pH of 5 (50% base saturation) and a CEC of $10\,\mathrm{cmol\,kg^{-1}}$ soil would be $10\,\mathrm{cmol\,kg^{-1}}$ soil $\times\, 40\% = 4\,\mathrm{cmol\,lime\,kg^{-1}}$ soil, which is equal to $4000\,\mathrm{kg\,lime\,ha^{-1}}$. One of the limitations of this approach is in estimating the CEC of the soil. If the estimate of the CEC is off, this will result in recommendations which are either too much or too little of the required amount of lime.

Titration of Soil with a Base

The lime requirement of soils can be determined by titrating acid soils to the desired pH with a base such as KOH or $Ca(OH)_2$. Titration of mineral soils to pH 5.6–5.8 will give the amount of lime required to neutralize exchangeable Al. The amount of lime on an equivalent basis is more than the exchangeable Al because, as the pH increases, nonexchangeable acidity ionizes and reacts with the lime. Titration to pH values above 5.6–5.8 results in the reactions of the base with nonexchangeable acidity which is H associated with hydroxy Al, carboxyl groups of organic matter, and hydrated oxides of Fe and Al.

Reaction of lime with exchangeable and nonexchangeable acidity is shown by the following equations:

1. neutralization of exchangeable acidity:

$$2\text{Al-soil} + 3\text{CaCO}_3 + 6\text{H}_2\text{O} \rightarrow 4\text{Ca-soil} + 2\text{Al(OH)}_3 + 3\text{H}_2\text{CO}_3$$

$$3\text{H}_2\text{CO}_3 \rightarrow 3\text{H}_2\text{O} + 3\text{CO}_2\uparrow \quad [3]$$

2. neutralization of nonexchangeable acidity:

$$2(\text{R-C-OOH}) + \text{CaCO}_3 \rightarrow \text{Ca(R-C-OO)}_2 + 2\text{H}_2\text{CO}_3$$

$$2\text{H}_2\text{CO}_3 \rightarrow 2\text{H}_2\text{O} + 2\text{CO}_2\uparrow \quad [4]$$

This method is rather time-consuming and does not lend itself to use where large number of samples need to be analyzed.

Buffer pH Method

Many soil-testing laboratories use the soil-buffer pH method to determine the lime requirement. This method is essentially a titration of an acid, the soil, with a base, the buffer solution. The molarity of the buffer solution is known and the amount of base required to neutralize the acidity of the soil can be determined by the change in the pH of the buffer solutions. With these methods a given volume of the buffer solution is added to a given weight or volume of soil and the pH is measured. The decrease in pH of the buffer solution is related to the amount of acidity in the soil.

Four buffer pH methods are being used in the USA to determine lime requirements. The amount of lime needed for each 0.1 decrease of buffer pH to bring the soil to the indicated pH is given in Table 5 for each of the methods. The Woodruff buffer was developed for determining the amount of lime to adjust the pH of Mollisols to a range of 6.5–7.0. The SMP buffer is used for soils with a large amount of three-layer clays and high-organic-matter content, such as Alfisols. The Adams and Evans buffer was developed for soils with low CEC and kaolinitic clay mineralogy. The lime recommendation by this method is to adjust the pH to 6.5. The Mehlich buffer was developed to predict the amount of lime needed to neutralize exchangeable Al. For mineral soils lime rates based on the acidity extracted with the Mehlich buffer raise the pH in the range of 5.8–6.0, which neutralizes the exchangeable Al.

Exchangeable Al

Liming mineral soils on an equivalent basis to the amount of exchangeable Al targets achievement of pH 5.6–5.8 and a soil solution Al^{3+} concentration approaching zero. As previously discussed, lime will neutralize exchangeable Al and also react with protons on variable-charge mineral surfaces. Consequently, the lime equivalence factor to achieve 0% Al saturation of the soil CEC is greater than 1.0 and is usually in the order of 1.5–3.

Yields for many crops are not reduced until soil Al saturation values exceed 40% or even 60% (Table 3 and Figure 5). For these acid-tolerant crops and, especially, in regions where lime materials are expensive, lime requirements can be adjusted to achieve a targeted percentage of Al saturation instead of neutralizing all of the exchangeable Al. An example of this approach is the following:

$$\text{CaCO}_3 \text{ equivalent } (\text{t ha}^{-1}) = \text{LF}[\text{Al} - \text{TAS}(\text{Al} + \text{Ca} + \text{Mg})/100] \quad [5]$$

where LF = the product of 1.5 equivalents of CaCO_3/equivalent of exchangeable Al and the conversion to a 20-cm layer for a hectare of soil, Al, Ca, and Mg = exchangeable cations extracted with a neutral unbuffered salt, such as KCl, and TAS = the targeted percentage of Al saturation after liming.

Liming soils to achieve a given percentage of Al saturation requires more information than when liming to neutralize all of the exchangeable Al. In addition to exchangeable Al, analytical data are needed for exchangeable Ca and Mg. Furthermore,

Table 5 Lime requirement as determined with buffer solutions

Method	*Buffer pH*	*Target pH*	*Intended use*	*Lime/0.1 pH decrease in buffer ($t\ CaCO_3\ ha^{-1}$)*
SMP	6.8	6.8	High exchangeable Al	0.6
		6.4	Alfisols	0.5
		6.0		0.4
Woodruff	7.0	6.5–7.0	Mollisols	0.5
Adams and Evans	8.0	6.5	Low cation exchange capacity Ultisols	0.1
Mehlich	6.6	6.0	Neutralize exchangeable Al, Ultisols	0.16

Data from van Lierop W (1990) Soil pH and lime requirement determination. In: Westerman RL (ed.) *Soil Testing and Plant Analysis*, 3rd edn, pp. 73–126. Madison, WI: Soil Science Society of America.

CEC, cation exchange capacity.

the percentage of Al saturation tolerated by the intended crop must also be known.

Factors to Consider in Lime Application

Liming Materials

There are a number of different materials which can be used for liming. The most common ones are calcium carbonate (calcitic limestone) and calcium-magnesium carbonate (dolomitic limestone). These are crystalline compounds which must be ground. Marls are soft, unconsolidated calcium carbonates which may contain some clay. Calcium oxide (burnt lime or quick lime) is a white powder and is caustic when it absorbs water. It is difficult to get uniform mixing with the soil because immediately after application absorption of water causes granules to form. Calcium hydroxide (hydrated lime) is a white powder and when wet is caustic. It is also difficult to apply. Basic slag (calcium silicate) is a by-product of the steel industry.

Neutralizing Value

The neutralizing value ($CaCO_3$ equivalent) of liming materials is related to the amount of acid that a unit weight of lime will neutralize. The standard for evaluating materials is pure calcium carbonate which is given a value of 100%. The neutralizing value for pure liming materials is given in Table 6.

The purity of the liming material must also be taken into account when determining the neutralizing value. Most liming materials are not 100% pure. In many instances the purity of the material will have been determined so that the buyer is given this information. Lime recommendations are often based on 100% purity of the material so that the application rate will have to be increased when the purity is less than 100%.

Fineness of Liming Materials

The most commonly used liming materials, calcitic and dolomitic limestone, are crystalline compounds. These materials must be ground to a fineness which will provide a large number of particles per unit volume of soil. The lime particles dissolve when coming into contact with acid soils and will release Ca^{2+} and HCO_3^-: this results in neutralization of H^+ and Al^{3+}. The rate at which the pH of the soil mass is affected depends on how close the individual lime particles are to one another. Since the diffusion of Ca is relatively slow the particles need to be close together so that the zones of neutralization around each particle overlap in a relatively short time.

Table 6 Neutralizing value of pure liming materials

Material	*Neutralizing value (%)*
Calcium carbonate, $CaCO_3$	100
Calcium-magnesium carbonate, $CaMg(CO_3)_2$	109
Calcium oxide, CaO	179
Calcium hydroxide, $Ca(OH)_2$	135
Calcium silicate, $CaSiO_3$	86

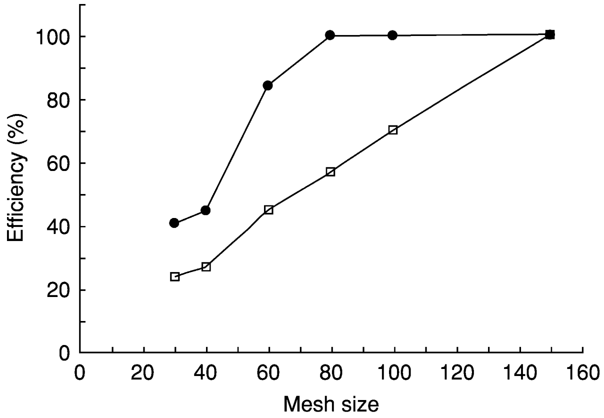


Figure 7 Relative efficiency of various mesh sizes of calcitic (circles) and dolomitic (squares) lime in neutralizing acidity. Data source: unpublished data of E.J. Kamprath.

Calcitic and dolomitic limestones are generally ground so that all of the particles pass a 20-mesh sieve and a certain percentage passes a 100-mesh sieve depending on the regulations of the state or province. Calcitic limestones are often softer than dolomitic limestones and therefore dissolve at a faster rate. The relative efficiency of different particle sizes of these two materials is given in Figure 7. Particles held on a 10-mesh sieve have little value for neutralizing acid soils.

Incorporation of Lime in the Soil

Since Ca^{+2} and HCO_3^- ions have a slow rate of diffusion, the lime must be well mixed with the soil to neutralize soil acidity throughout the zone which is to be amended. One approach is to apply half of the lime and incorporate it by disking, and then apply the other half and disk to incorporate. However this results in high costs of application and incorporation rather than doing it only once.

Conclusion

Poor plant growth on acid soils is due to toxicities of Al, H, and Mn and deficiencies of Ca, Mg, and Mo. Oxisols, Ultisols, and some Alfisols in their native state have pH values less than 5 and exchangeable Al of the effective CEC greater than 50%. Plant response to liming is due to neutralization of the toxic elements and the supply of Ca and Mg. The lime requirement for a soil can be based on the amount of lime required to: (1) adjust the pH to a desired value; (2) neutralize exchangeable Al; or (3) neutralize the acidity measured by buffer pH methods. Factors to consider in selecting a lime material are fineness, purity, and the need for Mg.

See also: **Acidity**; **Calcium and Magnesium in Soils**; **pH**

Further Reading

Adams F (ed.) (1984) *Soil Acidity and Liming*, 2nd edn. Madison, WI: American Society of Agronomy, Crop Science Society of America, Soil Science Society of America.

Alley MM (1981) Short-term chemical and crop yield responses to limestone applications. *Agronomy Journal* 73: 687–689.

Cochrane TT, Salinas JG, and Sanchez PA (1980) An equation for liming acid soil mineral soils to compensate crop aluminium tolerance. *Tropical Agriculture* 57: 133–140.

Evans CE (1968) *Ion Exchange Relationships of Aluminum and Calcium in Soils as Influenced by Organic Matter*. PhD thesis. Raleigh, NC: North Carolina State University.

Evans CE and Kamprath EJ (1970) Lime response as related to percent Al saturation, solution Al, and organic matter content. *Soil Science Society of America Proceedings* 34: 895–896.

Fox RH (1979) Soil pH, aluminum saturation and corn grain yield. *Soil Science* 127: 330–334.

Gonzalez-Erico E, Kamprath EJ, Naderman GC, and Soares WV (1979) Effect of depth of lime incorporation on the growth of corn on an Oxisol of Central Brazil. *Soil Science Society of America Journal* 43: 1155–1158.

Jallah JK (1994) *Assessment of Some Chemical Constraints to Root Growth in Four Acid North Carolina Soils*. PhD thesis. Raleigh, NC: North Carolina State University.

Kamprath EJ (1970) Exchangeable aluminum as a criterion for liming leached mineral soils. *Soil Science Society of America Proceedings* 34: 252–254.

McCart GD and Kamprath EJ (1965) Supplying calcium and magnesium for cotton on sandy, low cation exchange soils. *Agronomy Journal* 57: 404–406.

Osmond DL, Smyth TJ, Reid WS *et al.* (2002) *Nutrient Management Support System, NuMaSS*. (version 2.0). Raleigh, NC: Soil Management Collaborative Research Support System, North Carolina State University.

Ritchey KD, Silva JE, and Sousa DMG (1983) Relação entre teor de calcio no solo e desenvolvimento de raizes avaliado por um metodo biologico. *Revista Brasileira Ciencia Solo* 7: 269–275.

Sanzonowicz C, Smyth TJ, and Israel DW (1998) Calcium alleviation of hydrogen and aluminum inhibition of soybean root extension from limed soil into acid subsurface solutions. *Journal of Plant Nutrition* 21: 785–804.

van Lierop W (1990) Soil pH and lime requirement determination. In: Westerman RL (ed.) *Soil Testing and Plant Analysis*, 3rd edn, pp. 73–126. Madison, WI: Soil Science Society of America.

LIPMAN, JACOB G.

J C F Tedrow, Formerly with Rutgers University,
New Brunswick, NJ, USA

Jacob G. Lipman was one of the most prominent international leaders in soil science. In 1901 he formed the Department of Soil Chemistry and Bacteriology at Rutgers University, the first of its kind in the USA, and probably the world. He was the driving force in internationalizing soil science. He also initiated and edited Soil Science, the first international journal dealing with soil chemistry, soil fertility, soil physics, and soil microbiology.

Lipman (Figure 1) played a major role in the development of agriculture in the USA and beyond during the 1901–1939 period. His main contributions were in the area of soil microbiology and soil fertility, particularly in relation to increased crop production. During later years, he was also involved in worldwide agricultural problems. Not only did he advance science on several fronts but he was also a gifted administrator. He addressed the broader aspects of agriculture, particularly in increasing the world food supply, controlling soil erosion, and improving the general welfare of humanity.

The Early Years

Jacob Goodale Lipman was born in Friedrichstat, Russia, on November 18, 1874. He received his early academic training under private tutors in Moscow and in the classical gymnasium in Orenberg. In 1888, with his parents, he settled at the Baron de Hirsch colony in Woodbine, New Jersey. After graduating from farm school there he entered Rutgers

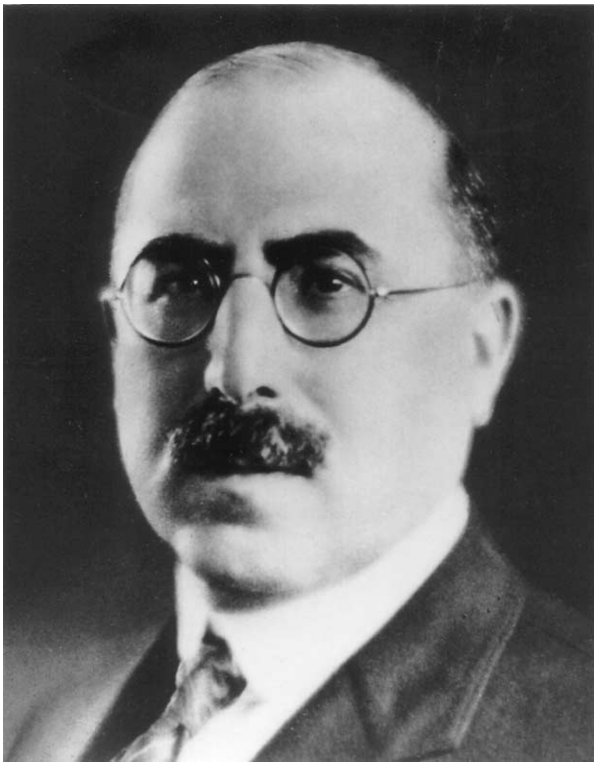


Figure 1 Jacob G. Lipman, 1874–1939.

College (University) in 1894. Next was graduate study at Cornell University where he studied under professors G.A. Caldwell and V.A. Moore. His master of arts degree was completed in 1900 and the doctor of philosophy degree in 1903. As his graduate studies were nearing completion he returned to the New Jersey Agricultural Experiment Station to organize a department of soil chemistry and bacteriology, the first of its kind in the USA. In 1902 he was appointed instructor in agriculture and by 1910 he had attained his (full) professorship. In 1911 he was appointed director of the Agricultural Experiment Station, but also retained his position as chairman of the Department of Soil Chemistry and Bacteriology. Additionally, in 1916 he was appointed as dean of the College of Agriculture, the first time that the office existed.

Lipman the Scientist

Early in his career people were just beginning to gain knowledge of the role of microbes and their applications to soil processes and to agriculture in general. During this time microbes were being recognized as agents responsible for numerous reactions in the soil that led to the liberation of chemical nutrients essential for plant growth.

Immediately after taking office, Lipman outlined a policy of increased service to the state. He saw the need for strengthening and enlarging the staff and adding new departments. This, however, could not be done without increased state appropriations.

During his first address to the Board of Managers after becoming director of the New Jersey Agricultural Experiment Station, Lipman announced the beginning of a policy of increased service to the state. He asserted that there should be a sound basis for development of the land. He also urged the expansion of extension activities. Through his close relationship with leading farmers and prominent businessmen, he convinced the legislature that there was a real need for greater financial support for agriculture. Accordingly, during his tenure, state appropriations were increased 16-fold. Legislators in Trenton looked forward to Lipman's appearance before the Appropriations Committee, not only for the humorous stories he told so well, but also because they thought he was sincere and they liked him. Pointing out that, for many years, the station had been a creative force in developing agriculture in New Jersey and beyond, he asserted that it was time to undertake a comprehensive campaign in the public interest. He proposed studies in farm management, irrigation, soil drainage, fertilization of crops, and product marketing. He also pointed out the need for upgrading the laws governing seed control, fertilizer manufacture, and animal feeds.

Dr. Lipman's own scientific activities began at the turn of the century when the importance of microorganisms in soil processes was beginning to receive increased attention, particularly relating to nitrogen studies. In the history of soil microbiology, the name of Lipman will forever be connected with the study of the cycle of nitrogen in nature, especially the problem of nitrogen fixation by nonsymbiotic bacteria, the liberation of nitrogen as ammonia in the process of decomposition of nitrogenous materials in the soil, and the general relations of nitrogen liberation in soil to the growth of vascular plants.

One of Dr. Lipman's most important scientific contributions dealt with the transformation and fixation of nitrogen by bacteria. In studying some of the organisms concerned in the process he broached two subjects that later developed greatly: the importance of joint action of the soil organisms, and the relation between denitrification and composition of the organic matter concerned, now commonly referred to as the carbon–nitrogen ratio.

During his studies Lipman isolated a newly discovered nonsymbiotic nitrogen-fixing organism, *Azotobacter vinelandii*; he also described several other newly discovered organisms. The genus Azotobacter was recognized as one of the important members of the soil population. Further, Lipman was one of the first to prove that Azotobacter are capable of fixing large quantities of nitrogen from the atmosphere. Later, he identified two more important species of Azotobacter: *A. beijerinckii* and *A. woodstownii*. Lipman showed that the same soil type under different farming systems had the ability to produce ammonia from various organic compounds. Further, Lipman's work can be considered the forerunner of the Azotobacter method for determining the presence and abundance of available minerals in the soil.

The next signal accomplishment of Dr. Lipman in soil research concerned oxidation of sulfur in the soil. The underlying task was to make rock phosphate more available to plants. Lipman believed that if certain bacteria in the soil were capable of oxidizing the sulfur, the phosphate would be converted into a more soluble form. Together with his assistants, Selman A. Waksman and Jacob S. Joffe, he isolated the bacterium *Thiobacillus thiooxidans* and applied it to the soil. The concept had considerable agricultural appeal and patents were granted covering the process. Subsequent investigations by Lipman's colleagues, however, showed that in New Jersey soils the rate of sulfur oxidation in the uninoculated samples was about the same as that of the inoculated ones.

Dr. Lipman's scientific accomplishments soon achieved worldwide recognition. Students came from many foreign lands to carry out their graduate studies at Rutgers University. In all, students from some 24 foreign countries studied under Lipman and colleagues. Among the future luminaries from the USA and beyond were P.E. Brown, R.T. Dubos, J.S. Joffe, W. Rudolfs, R.L. Starkey, and S.A. Waksman. Further, young investigators who had already completed their doctor of philosophy degrees, as exemplified by Hans Jenny and W.L. Kubiëna, worked in the department for shorter periods.

It was often said that Dr. Lipman kept his associates happy and smiling. He believed in delegating responsibility and in so doing inspired his associates to supreme efforts. But once he delegated an assignment he would hold his staff member responsible. One staff member commented that Lipman could outline in one-half hour more than could be done in 6 months. His colleagues also said of him that, while demanding of students, he was also patient and understanding. He was never known to punish anyone. He felt that everyone should have another chance if they wanted it and he was never known to be angry with any subordinate.

As a writer and speaker Dr. Lipman was at his best. He delivered a series of lectures at other institutions such as University of Illinois, Cornell University, University of Tennessee, University of Nebraska, and Iowa State University. He was also called upon to speak at many agricultural meetings in New Jersey and beyond.

After he became director of the New Jersey Agricultural Experiment Station visitors came to learn from him, not only from nearby but also from all over the world and from all walks of life. Lipman had an encyclopedic mind. His conversations were not limited to agriculture but included general science, economics, and religion. Very few investigators have the skills to master the field of research and also have the ability to convey messages to the ordinary farmer. Lipman, however, bridged the gap admirably. He wrote many reports on the technical phases of soil science and agriculture and contributed articles to various farm journals such as Pennsylvania Farmer, Rural New Yorker, New Jersey Agriculture, and the John Wiley Agricultural Science series. He wrote technical reports for C.E. Marshall's Microbiology and E.O. Jordan and I.S. Falk's book The Newer Knowledge of Bacteriology and Immunology. Lipman was also an advisory editor of the Journal of Agricultural Research, the Internationale Mitteilungen für Bodenkunde, and Annales Agronomiques. Dr. Lipman also edited a series of textbooks in agricultural science. His Bacteria in Relation to Country Life gave him nationwide recognition. This was followed by his Laboratory Guide of Soil Bacteriology, coauthored with his first assistant, P.E. Brown.

Dr. Lipman held membership in a large number of professional societies throughout the USA. The list included the American Health Association, American Association for the Advancement of Science, American Chemical Society, Society for the Promotion of Agricultural Science, American Society of Agronomy, Association of Official Agricultural Chemists, Society of American Bacteriologists, Association of Land Grant Colleges, New Jersey State Sanitary Association, Washington Academy of Science, American Academy of Political and Social Science, National Research Council, American Peat Society, New Jersey Chemical Society, and others. The list of his membership in foreign societies included the International Commission of Agriculture Ecology, Reale Accademia del Georgofili de Firenze, French Academy of Agriculture, Academy of Agriculture of Czechoslovakia, and the Swedish Royal Academy of Agriculture. In 1922 he was awarded a silver medal by the French Academy of Agriculture. Through Lipman's efforts the New Jersey Agricultural Experiment Station received the Bacteria Award at the Panama-Pacific International Exposition of 1915.

In 1923 Rutgers University awarded Dr. Lipman an honorary doctor of science degree and in 1930 he received an honorary doctor degree from the Catholic University of Santiago, Chile.

Dr. Lipman's research subjects during the 1920s and 1930s included soil microbiological populations, the influence of environmental conditions upon biological changes in the soil, the correlation of soil fertility and microbiological processes, enzymes of microorganisms, anaerobic fermentation of hemicelluloses, and transformation of organic materials.

Dr. Lipman's investigations were carried out during a period of soil microbiology when both scientists and practical microbiologists were seeking applications of science to agriculture. His work was carried out during the time when there was a transition between science developed in the laboratory and application of the results to field conditions. He contributed to both and thus carried the service through one of the most complex periods of development. He not only helped to advance science of the soil, but he also interpreted the application of science to practical agriculture.

At the height of his career Dr. Lipman turned to the broader issues of society. He was one of the pioneers in advancing rural settlement, rural rehabilitation, soil conservation, flood control, afforestation, wildlife sanctuaries, and the general good use of land.

Up until the World War I era there were virtually no important technical journals dealing specifically with soil science published in North America. Americans had to rely mainly on technical papers and books dealing with soil science from Germany and England. The great body of soils information from Russia had not, at that time, penetrated the western world to any extent. In 1916 Lipman founded and edited Soil Science, a journal he personally continued editing for the remainder of his lifetime. In all he edited 46 volumes (two volumes per year). Contributions dealing with soil science came from investigators from all over the globe. Lipman, at an early date, appointed some 35 consulting editors of the journal, over half of whom were from foreign lands. Soil Science was truly an international journal without any society or government affiliation. This venerable journal, published monthly since its inception, has had, over the years, 2000–3000 institutional and private subscribers, with nearly half of them from foreign lands. Lipman's labors with the journal alone served as a vehicle for internationalizing the field of soil science.

At the time when Dr. Lipman became dean of the College of Agriculture and director of the Agricultural Experiment Station, the soils of New Jersey were being surveyed on a scale of 1 in per mile (2.5 cm per 1.6 km). The surveys were being carried out as a joint project with the US Department of Agriculture. By 1927 surveys for the entire state were completed. New Jersey was one of the first states to have such a survey. Soils were also collected and analyzed for their chemical, physical, and mineralogical properties. Also, there were recommendations for cropping systems, land management, soil drainage, and related subjects.

During the latter part of his life Dr. Lipman made a number of trips to Europe, particularly England, the Netherlands, Germany, France, and Italy. Conversing with non-English-speaking colleagues presented few problems because he was fluent in Russian, German, and French. European investigators returned many visits to Lipman. He also made trips to Central and South America. His main purpose was to study various agricultural practices.

Lipman the Administrator

When Lipman became director of the New Jersey Agricultural Experiment Station the physical facilities at the station were rather limited. Office and laboratory space was limited to two buildings on the main campus of the university. There were also a few farm buildings at College Farm on the edge of New Brunswick. Lipman increased the acreage of College Farm from 142 to 800. Over a period of 20 years the entire operation of the Agricultural Experiment Station plus the teaching facilities in agriculture were moved to College Farm where they remain to this day.

During the 1911–1930 period many new, major structures plus ancillary buildings, including animal barns, were constructed at the College of Agriculture. By 1928, with the growth of research and the increase in the number of students, the state legislature approved the construction of an additional building to be devoted primarily to research and the teaching of soil science and in 1931 funds were provided for the new building. But just as the architectural plans were completed, the state, because of the world depression, rescinded all appropriations for capital construction. In 1950, however, the new building was completed and it bears the name Lipman Hall (Figure 2).

At the time when Dr. Lipman became director of the New Jersey Agricultural Experiment Station there were only a few departments. In addition to enhancing the physical plant he created departments of seed analysis, plant pathology, agronomy, and farm management. During the later years he also added departments of plant physiology, agricultural biochemistry, agricultural engineering, sewage research, and ornamental horticulture.

Dr. Lipman greatly increased the size and scope of the New Jersey Agricultural Experiment Station but there were times when it was difficult to obtain sufficient funds to carry out his plans, particularly during the World War I era. He was heard to say: "I wake up

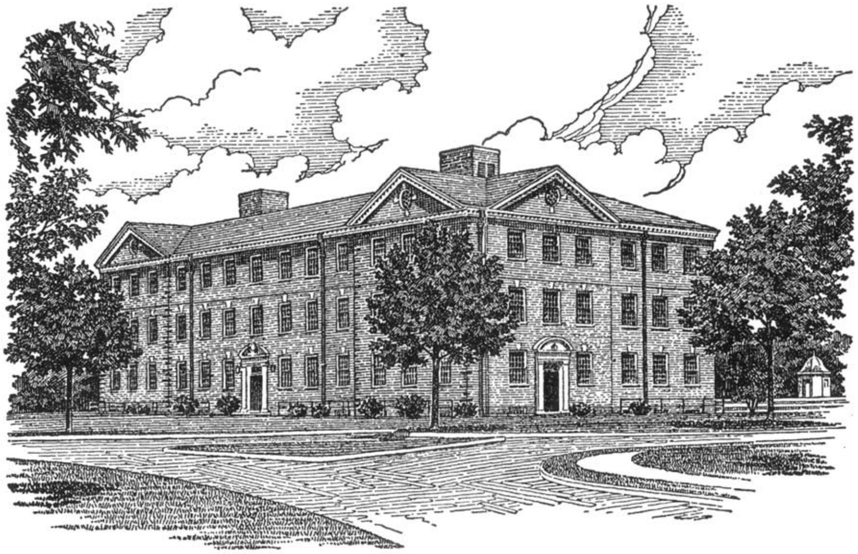


Figure 2 Lipman Hall on the campus of Rutgers University, New Brunswick.

Figure 3 Experimental soil cylinders at the New Jersey Agricultural Experiment Station, Rutgers University, New Brunswick.

in the morning with a vision of a pile of mail waiting for me, requests I cannot fill, complaints I cannot help, and work piled without end. I have an idea I'd rather prefer the trenches."

Dr. Lipman was well aware of the many varieties of soils in New Jersey. Accordingly, he set up management demonstrations on many farms. Branch stations were established for specialized crops such as blueberries and cranberries. Also, a shellfish research station was established, as was a dairy research station.

Experimental work dealing with growing crops on different kinds of soil was initiated by Dr. Lipman's predecessor, Dr. E.B. Voorhees. Initially there were 60 openend, steel cylinders, 4 ft (1.2 m) high with a diameter of 23 1/2 in. (60 cm) installed in a cluster at ground level. Various soils were placed in the cylinders and crops such as tomatoes, corn, and oats were grown (Figure 3). Later, under Lipman's directorship, the number of cylinders was increased to 320 and additional experiments included various soils and fertilizer mixtures. Also, chemical and bacteriological studies were made regarding the pathways of nitrogen and organic matter transformation. In addition to gaining general information on nutrient balance the cylinder experiments demonstrated that enormous losses of nitrogen may take place in the soil. Also, it was found that soluble nitrogen in liquid manure possesses high value as a source of nitrogen. The cylinder experiments settled many worldwide problems dealing with the availability and fate of many kinds of fertilizers.

Lipman's predecessor at the State Agricultural Extension Service had been operating in New Jersey since about 1890 but the service was greatly expanded under Lipman. Early in his tenure as director there was a major consolidation of the various agricultural organizations in New Jersey. In 1912 extension activities were formally initiated. When the state legislature convened in 1913, New Jersey governor Woodrow Wilson had a bill drafted for an extension system of the governor's own design. Farm demonstrations were to be operated by county agents who would be responsible to a state superintendent appointed by, and responsible to, the governor. Wilson's idea was to have the extension service completely independent of the Agricultural Experiment Station. As soon as the draft bill was made public, Lipman called on the governor and urged that the extension service be incorporated into the college and station. But the governor said, "I want it my way." The bill passed the assembly but the governor was later due for a surprise. Lipman turned to his long-time friend Harry Wheaton, an elderly blacksmith who was chairman of the State Committee on Agriculture and urged him to "sit on the bill." March came and Wilson was inaugurated president. The controversy between Lipman and Woodrow Wilson continued but Lipman had convinced Wheaton to continue to withhold the bill from a final vote. Lipman said to him: "Woodrow Wilson is no longer Governor of New Jersey, now he is only President of the United States." Lipman persisted and, finally, had his own bill passed providing for the extension service to administer activities through Rutgers University.

Dr. Lipman realized the need for international collaboration particularly between investigators from Europe and those from the USA. During the Third International Conference on Soil Science in Prague in 1922 the USA delegated Lipman and C.F. Marbut (Head of the US Soil Survey) to attend. Lipman's mind had developed so keenly and rapidly as to extend well beyond the immediate problems with which soil science was struggling. At the Fourth Conference on Soil Science in Rome in 1924 Lipman again attended the meetings and continued to make very favorable impressions on European colleagues. It was during this meeting that the idea of forming the international society of soil science was finalized. Washington, DC, was selected as the meeting place and Lipman, because of his great organizational talent, was selected as president. Selection of the USA was based on several factors: (1) European investigators were anxious to make observations on US soils in general; (2) there was particular interest in seeing the highly weathered, reddish soils in the south-eastern USA; and (3) investigators wished to make studies and observe the similarity of the dark-colored soils of the prairies and steppes of the central USA to their counterparts in Eurasia.

The First International Society of Soil Science meeting was held during June 1927. President Calvin Coolidge gave the keynote address, which was followed by an address by Lipman. During the meeting Dr. Lipman said:

> Truly in man's upward climb he has accomplished the seemingly impossible. But much as the student of soils may be tempted to ponder on man's past, he must leave this to the philosopher and the poet. It is his task, rather, to inquire into the relations of climates, soils, plants, animals, and man. For all that, soil science must build a foundation large enough and strong enough to support the study of plant food resources, their mobilization, and the interrelations of soils and plants and the characteristics and peculiarities as reflected in the make-up of plants, animals, and man.

Immediately following the First International Soil Science Congress, Lipman was instrumental in arranging a 30-day, 12 000-mile (19 320-km) trip to the west coast of the USA and western Canada. A special train was engaged and it carried over 200 people in 17 railroad cars. The excursion gave delegates the opportunity to make firsthand observations concerning North American soils and make comparisons to European conditions. One individual went so far as to say that the transcontinental excursion was as important to his career as the voyage of the Beagle had been to Charles Darwin.

The End of an Era

In February 1939 Dr. Lipman developed a severe cold and was confined to his home. He was hospitalized several times during the following weeks. Local physicians found his heart completely gone. He died on April 19, 1939. His services to humanity, however, were not forgotten, as exemplified by the following tributes made by various speakers during the dedication of Lipman Hall.

Nobel Laureate Selman A. Waksman said that, with the death of Dr. Lipman, Rutgers completed a period of nearly a century of investigation in the field of soil science, a science which, even if not touched by this group and their coworkers in all its phases, has reached some of the highest pinnacles of its attainment. The investigations of the soil, its origin and nature, its fertility and its relation to plant growth, its microscopic population, and its microbiological processes are closely interwoven with the other investigations carried out not only in the various institutions in this country but in the world at large. They have served to raise the science of the soil from its beginnings, as a child of geology and as a handmaiden of agronomy to an independent, natural science. We shall provide then for the intensification of production and the support in greater comfort of the growing population of the world. It is the duty of soil sciences to establish more clearly the relation of crops to soils in order that cropping methods may best conserve plant food, solar energy, and human labor. Soil science should teach the farmers how best to modify food quality by soil treatment. It should plan a more far-reaching program of research and education in order that it might serve in a constantly larger way the needs of a human society that from year to year is growing more complex in its organization and more exacting in its demand for such of the products of the land as are essential for maintaining the moral and spiritual values that are the final measure of human activities.

Sir E.J. Russell of England said that there are two ways in which a man may advance science: by his own personal investigations and by organizing investigations of others. And, since nothing succeeds like success, he must ensure success by keeping the whole program well-balanced, favoring none, repressing none. Dr. Lipman followed the above usual sequence and he was successful in all parts he had been called upon to play. His long and faithful service to the New Jersey Agricultural Experiment Station can never be forgotten, nor can his pioneer work in opening up new branches of knowledge.

Further, Rutgers president R.L. Clothier said of Lipman that he saw him as a man endowed with uncompromising intellectual honesty, relentless mental discipline, inspired reasoning, dogged perseverance, and a rare vision of the future of pure and applied research. Clothier also found Lipman to be endowed with the qualities of character and personality that made him a sympathetic friend, a rare raconteur, and a delightful companion.

See also: **Jenny, Hans**; **Marbut, Curtis Fletcher**; **Waksman, Selman A.**

Further Reading

Allison RV (1935) Jacob G Lipman as teacher and Director of Research. *Soil Science* 40: 31–36.

Blair AW (1935) Jacob G Lipman and New Jersey Agriculture. *Soil Science* 40: 25–29.

Martin WH (1952) *Lipman Hall*. New Brunswick: Rutgers University.

Russell EJ (1935) Jacob G Lipman and Soil Science. *Soil Science* 40: 3–7.

Waksman SA (1935) Jacob G Lipman as an investigator. *Soil Science* 40: 11–25.

Waksman SA (1966) *Jacob G Lipman: Agricultural Scientist and Humanitarian*. New Brunswick: Rutgers University Press.

Woodward CR and Waller IN (1952) *New Jersey's Agricultural Experiment Station 1880–1930*. New Brunswick: Rutgers University.

LOESS

A J Busacca and M R Sweeney, Washington State University, Pullman, WA, USA

Introduction

Loess is a terrestrial deposit of eolian (wind-blown) dust composed predominantly of silt-sized particles. Most loess has been altered weakly to strongly by soil-forming processes as it has accumulated and its unique properties create highly fertile agricultural soils that are some of the most productive in the world. Loess covers large land areas, most often downwind of the meltwater streams of large Pleistocene glaciers, although loess is also found downwind of some large and small desert areas. Erosion of loess soils via wind and water can be severe and reduces crop productivity; however, the global redistribution of small amounts of fine dust by wind to areas of highly weathered and leached soils has recently been recognized as adding nutrients that allow these soils to sustain complex ecosystems throughout the world. Buried soil sequences within thick deposits of loess can be used to reconstruct the history of climate and ecological changes during the Pleistocene and Holocene Epochs, covering approximately the last 2.4 Ma of earth history.

Distribution

Loess covers approximately 10% of the earth's land surface (Figure 1) and is generally associated with semiarid to semihumid climate regimes. The thickest deposits are found in China, where approximately 500 000 km^2 is blanketed by loess that is up to 300 m thick. The Great Plains has the largest area of loess in North America. Loess there can be more than 60 m thick. Other loess deposits include the Pampas of Argentina, parts of Europe, the Ukraine, large areas in central Asia, the Palouse and Snake River Plain of the US Pacific Northwest (PNW US), parts of Alaska, and parts of northern Africa. Thinner and less extensive deposits are found worldwide.

Origin of Silt and Eolian Transport of Loess

There are three main processes responsible for the formation of loess: (1) production of silt, (2) wind sufficient to entrain the silt into suspension in the atmosphere, and (3) suitable trap areas where loess can accumulate owing to surface soil moisture or vegetative cover.

Glacial grinding is a dominant producer of fine-grained sediment; however, silt is also produced by fluvial and eolian abrasion, frost-shattering, salt-weathering, chemical-weathering, biological processes, and tectonic processes (i.e., overburden release). Since loess is sometimes found downwind of dune fields, it has been argued that grain-to-grain collision during saltation of sand creates silt particles, but this is only a minor contributor to dust production. Silt, produced by one of the above mechanisms, is commonly transported from its place of origin and redeposited before it is entrained by wind to form loess. Models for different types of loess help to explain its genesis. High-latitude loess is the most extensive and includes periglacial and perimontane loess. Periglacial loess is derived from silts produced by glacial grinding that are transported by water on to outwash or braidplains that are subject to redistribution by wind erosion. Perimontane loess is associated with high rates of erosion in mountainous areas that supply silt for adjacent loess accumulations. More recently recognized, peridesert or 'desert' loess is derived from the eolian reworking of sediments from alluvial fans, dry lake beds, or fluvial channels in deserts. Loess accumulates at the semiarid margins of the desert, where an increase in vegetative cover acts as a trap for the silt. Peridesert loess deposits tend to be thinner, sandier, and less continuous compared to high-latitude loess.

Mineral particles can be entrained by the wind by several processes. Particles can be entrained when a threshold velocity is exceeded, resulting in aerodynamic lift of the particle, or by fluid drag exerted on

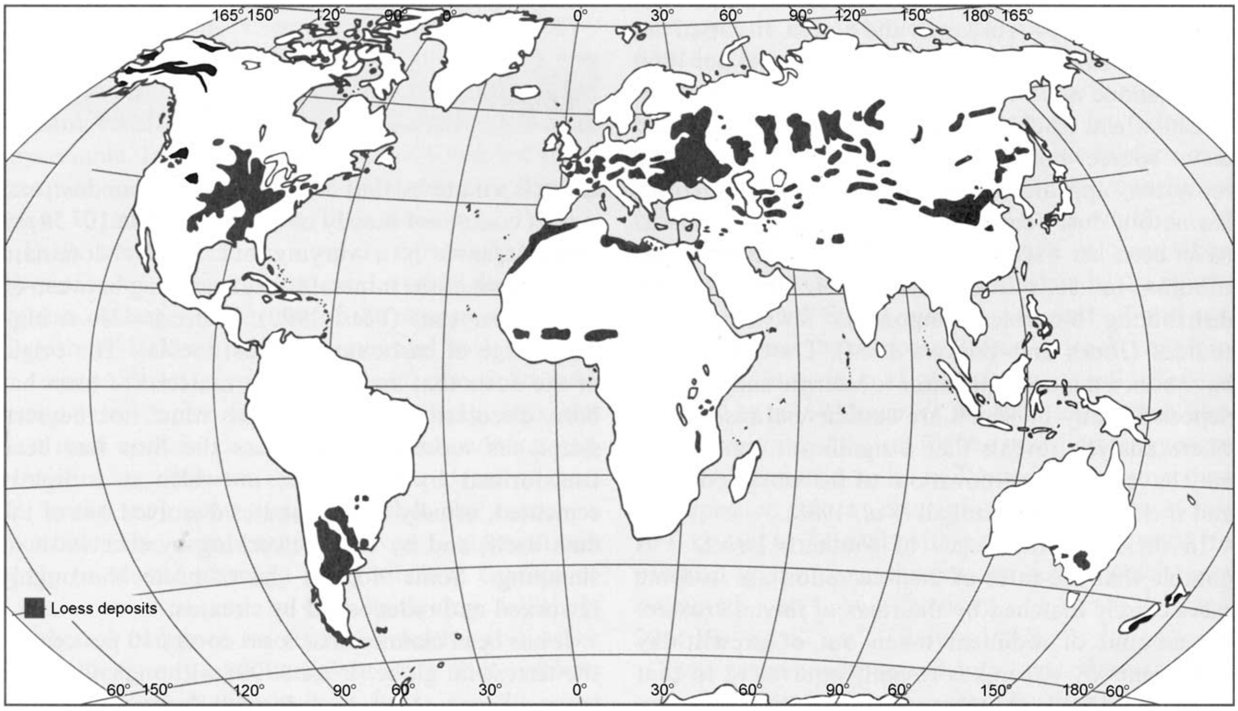


Figure 1 World distribution of major loess deposits. Adapted from AEOLIAN GEOMORPHOLOGY by Livingstone/Warren. © Reprinted by permission of Pearson Education, Inc., Upper Saddle River, NJ.

the particle by air. Particles may also be ejected into the air when they are struck by sand-sized grains. Silt- and clay-sized particles have electrostatic binding forces that can cause soil crusting and aggregation and inhibit their entrainment by the wind. (*See* **Erosion:** Wind-Induced.) For this reason, silt and clay are generally entrained as sand-sized aggregates or ejected by grain impacts. Entrainment is hindered by surface crusting, surface armoring (i.e., gravel lag), and surface soil moisture. Disturbance by animals and humans may aid in the deflation of dust.

Once silt and clay are released into the air column, transport is dominantly by suspension. While coarser sediments such as sand are transported close to the bed via creep (rolling) or saltation (jumping), the smaller silt and clay particles are carried vertically higher into the air column by turbulent eddies. Coarse silt and very fine sand are carried in short-term suspension and are deposited within tens of kilometers of the source. Finer silts and clays are carried in long-term suspension for tens to hundreds of kilometers or farther.

Deposition of dust particles out of suspension to form loess occurs when: (1) wind velocity and turbulence decreases; (2) particles are trapped by vegetation, moist soil surfaces, or uneven topography; or (3) particles are washed out of the atmosphere in precipitation events. If dust is deposited on a bare, dry surface, it is likely to be reentrained by the wind, because there is no means to hold it in place.

Sedimentology

Loess is composed predominantly of particles that range in size from 20 to 60 μm. Loess usually has a unimodal grain size distribution, is poorly sorted, and is finely skewed. Loess with more than 20% sand-sized particles can be called ‘sandy loess’ and that with more than 20% clay ‘clayey loess.’ Grain size of any loess is a function of the proximity to the dust source and the degree of weathering during soil formation. Loess near its source contains higher percentages of sand and coarse silt. Mean grain size typically decreases downwind, with a concomitant increase in fine silt and clay. This trend is a result of sorting of particles by the wind and decrease in wind velocities downwind. In some loess areas such as the Palouse of the PNW US, downwind sorting and fining are accompanied by increases in density of native vegetation and increased precipitation and surface soil moisture.

The color of loess ranges widely, depending on source mineralogy and weathering; it can be light to dark gray, yellow, pale to dark brown, red, or black. (*See* **Morphology.**) Mineralogy of loess reflects the source from which it is derived. Generally, loess is composed predominantly of quartz grains (50–70%) with feldspars, micas, carbonates, clay minerals, heavy minerals, and volcanic tephra. (*See* **Minerals, Primary.**) Clay minerals may be detrital (inherited from the source rocks) or formed *in situ* by weathering during soil formation. Carbonate minerals are often present in dust and may comprise more than 30% of loess. Inherited carbonate grains may be dissolved and precipitated to form pedogenic calcic horizons in semiarid to arid climates.

On a macroscopic scale, loess generally lacks sedimentary structures such as bedding and thus has a homogeneous appearance. Soil horizons can give exposures of loess a pseudobedded appearance. Frequent wetting and drying may result in formation of prismatic structure in loess. Closer detail of loess sedimentology may reveal fluctuations in the percentage of sand, silt, and clay with depth in the deposit that may reflect changes in wind intensity or in source areas through time. Rarely, laminations of coarse and fine dust have been preserved in loess that are interpreted to represent seasonal fluctuations in wind strength during times of very high dust accumulation rates that outpaced the mixing processes of soil organisms.

Stratigraphy

Loess accumulates in layers or ‘strata’ and soils commonly form at the top of a loess stratum. The soils become buried with continued deposition of dust, creating ‘stratigraphic sequences’ of alternating loess and buried soils (paleosols). Loess sequences such as those in China, the US, and in many other areas have been utilized to reconstruct the history of the Quaternary Period. The common occurrence of loess-paleosol sequences has been interpreted to represent episodes of high rates of dust accumulation during which a loess layer is deposited, alternating with episodes of low rates of dust deposition during which soils develop. Micromorphological techniques applied to loess and paleosols (microscope studies of slides of undisturbed materials) have shown that the idea of sequences recording either periods of dust accumulation or soil formation is misleading. In fact, soil-forming processes occur throughout the deposition of dust, and accumulation rates and climate affect the intensity of soil-forming processes (Figure 2).

Correlation of loess sequences across regions has been hampered by the lack of reliable dating techniques. Luminescence dating is a relatively new radiometric technique that dates the last exposure of the dust grains to sunlight. It can provide ages up to approximately 100 000 years or older with analytical

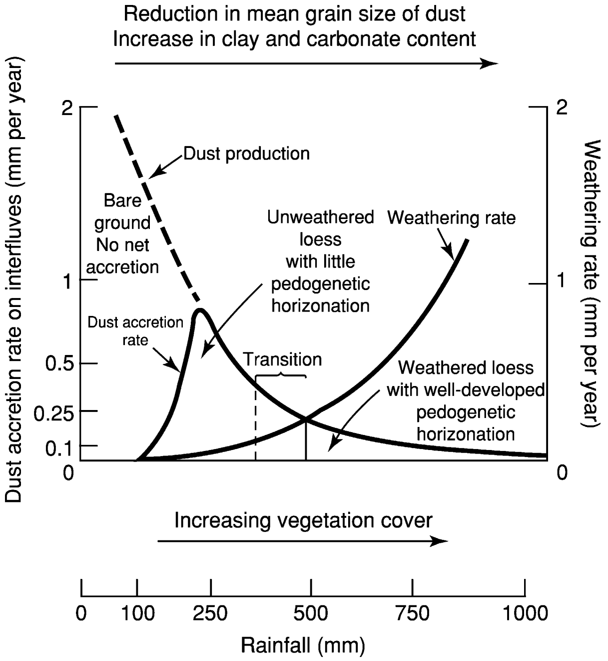


Figure 2 Schematic model showing increase in pedogenesis of loess with an increase in moisture and vegetation and decrease in dust accumulation rate and grain size. Reproduced with permission from Pye K and Tsoar H (1987) The mechanics and geological implications of dust transport and deposition in deserts with particular reference to loess formation and dune sand diagenesis in the northern Negev, Israel. In: Frostick L and Reid I (eds) *Desert Sediments: Ancient and Modern*, pp. 139–156. Special Publication No. 35. Geological Society.

errors as low as 5%. In some cases, prominent paleosols can be used to correlate sequences. Volcanic tephras in loess can be traced over large areas and can have unique chemical and mineralogical fingerprints that aid in correlation.

Chinese loess is as old as 7 million years, but the bulk of the loess there is less than 2.4 million years old. Loess deposition in China has been linked to the uplift of the Tibetan Plateau that may have resulted in climates more favorable to loess formation. Cycles of accumulation of loess followed by soil formation have been attributed in China to variations in the strength of the summer (rain bearing) compared with winter (dust-storm bearing) monsoons, and these have been correlated in turn to Milankovitch cycles (cycles in Earth's orbital parameters) that are thought by some to be the driver of the ice ages. Loess in the Great Plains and Europe show cycles that follow Pleistocene glaciations, with major loess accumulation during the dry and windy glacial periods, followed by soil formation during the more humid and less dusty interglacials.

Early scientists who studied loess were emphatic that loess was purely a glacial phenomenon. Some more recent studies show that there are cases where the opposite has occurred. Study of the Palouse loess indicates that maximum rates of loess deposition occurred during interglacials, while soil formation was dominant during glacials. The majority of the silt is glaciogenic, but it was transported to basins via catastrophic outburst floods triggered by ice dam failures that occurred toward the end of full glacial periods. The flood sediment was reworked by the wind and deposited as loess during interglacials. When full glacial conditions returned to the area, dust-transporting winds appear to have been suppressed by changes in atmospheric circulation, landscapes ceased aggrading, and soil formation ensued.

Geomorphology

Loess is a blanket deposit that accumulates from suspension fallout of dust particles. Regional patterns of thickness are controlled by a number of factors. Loess strata are thickest closest to the source area, as large volumes of the coarser sediment falls from suspension, leaving smaller volumes of finer sediment to be blown farther downwind. An exponential decrease in loess layer thickness with distance along downwind transects has been documented in many places such as the North American Great Plains (Figure 3). Thickness may be increased locally by the presence of trapping mechanisms such as topography or vegetation that separate the saltation load (sand that keeps the dust entrained) from the suspension load. A persistent source of dust-sized particles for eolian entrainment will also result in thick deposits and continuously aggrading land surfaces. The topographic form of loess landscapes can be quite different from place to place. The topography of the Palouse in the PNW US is a classic example. Loess in the drier, western Palouse (mean annual

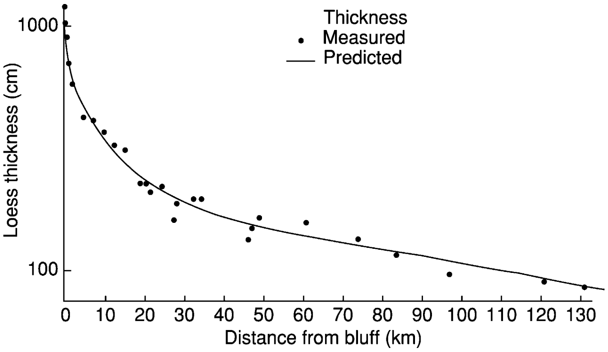


Figure 3 Exponential decrease in loess thickness downwind of the Missouri River source in Illinois, USA. Reproduced with permission from Frazee CJ, Fehrenbacher JB, and Krumbein WC (1970) Loess distribution from a source. *Soil Science Society of America Proceedings* 34: 296–301.

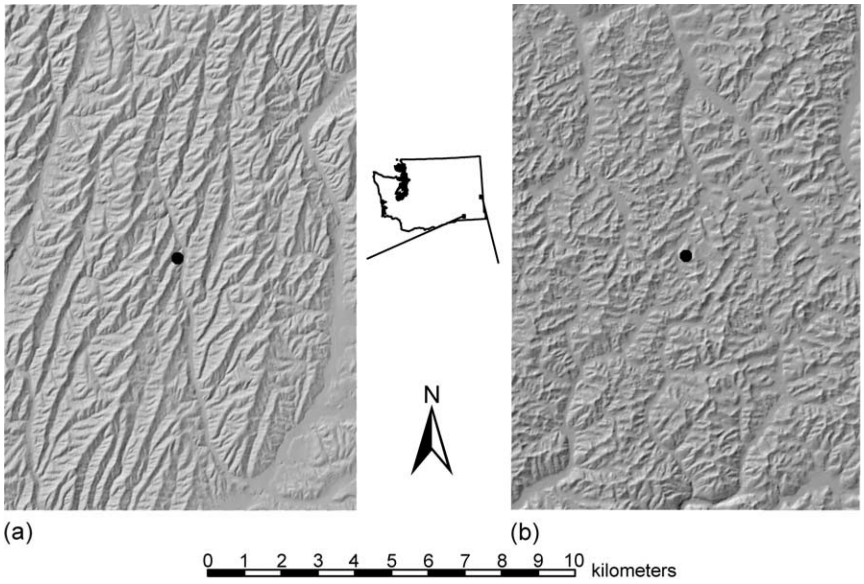


Figure 4 Hillshade models of Palouse topography, in Washington state (silhouette) in figure, based on 30-m digital elevation model resolution. × 2 vertical exaggeration. (a) Distinct, wind-aligned NNE linear ridges, formed during the deposition of sandy loess. Image of Hadley 7.5′ quadrangle; centroid is 118.44° west, 46.19° north; (b) rolling terrain of loess hills shaped predominantly by water erosion in the higher precipitation zone of the Palouse. Image of Garfield 7.5′ quadrangle; centroid is 117.19° west, 47.06° north.

rainfall less than 450 mm) forms a distinctive landscape of NNE-trending linear ridges and valleys (Figure 4a). This linear topography, called ‘greda’ or ‘paha,’ has also been observed in the loess of Hungary and the North American Great Plains. Linear ridges commonly form where loess has a more sandy texture. Local topographic highs may influence the deposition of loess in the lee of any preexisting topographic obstacles, resulting in streamlined ridges aligned with the prevailing wind direction. Other workers have argued that loess will accumulate on the windward side of topographic highs owing to deceleration of the wind, but this only may occur in the absence of sand particles. Loess in the wetter eastern Palouse (mean annual rainfall up to 800 mm) forms a dissected terrain of rolling hills with ‘cirque-like’ amphitheaters and a dendritic drainage network (Figure 4b). There, the combination of greater soil moisture, greater incidence and severity of soil freezing, and existence of buried soils with water-restricting, clayey argillic horizons create a landscape shaped by water erosion on frozen soils and rotational slumps and mass movements.

Loess can also blanket preexisting landforms. In eastern Europe, loess commonly blankets flights of fluvial terraces. Pseudoanticlines are described where loess has accumulated on undulating topography (Figure 5). Where dust accumulates in an area of higher relief, loess thickness can be greater in the valleys due to colluvial or alluvial reworking of loess off the hillslopes.

Figure 5 Thick accumulation of Palouse loess representing approximately 100 ky of deposition. Note the ‘pseudoanticline’ and the presence of buried soils represented by more resistant, white petrocalcic horizons. Photo by M. Sweeney.

Paleosols in Loess and Studies of Paleoclimate

Loess is considered to be one of the better terrestrial proxies for reconstructing paleoclimate for the Quaternary Period (the last 2.4 million years). This is at least in part because loess deposits aggrade through time, so that the land surface rises as dust accumulates and aggradation results in the preservation of former surface soils, or paleosols.

When soil formation keeps pace with dust deposition, soil horizons such as calcic, cambic, or argillic horizons form upward over time within the accumulating loess. In situations where loess deposition and soil formation are more or less continuous, buried soils may be indistinct and hard to recognize, because subsoil (B horizon) features are superimposed on topsoil (A horizon) features. 'Welded soils' occur where increments of loess are insufficiently thick to separate episodes of soil development completely, resulting in partial overlap of soil horizons and modification of the properties of buried soils. Only where episodes of dust deposition are episodic and new loess layers are thicker than typical soil profiles are buried soils with distinct A–B–C horizon sequences likely to be preserved.

There are several properties of paleosols that can provide clues to past climates including soil morphology, stable isotopes in pedogenic carbonates and soil organic matter, magnetic susceptibility, preservation of plant opal phytoliths, and plant pollen (Figure 6).

The formation of a soil in loess itself is very informative, as it implies a period of relative landscape stability and decreased dust deposition. Specific soil features can be key indicators of past climates. For example, calcic (Bk), petrocalcic (Bkm), or duripan (Bqm) horizons suggest arid climates, whereas argillic (Bt) horizons may indicate higher precipitation that drove clay illuviation processes, and redoximorphic features suggest changing soil moisture conditions with periodic anaerobism. (*See* **Morphology**.) Platy soil structure may be indicative of freezing conditions. Faunal fabrics formed by burrowing organisms have also provided clues to past ecological conditions. Palouse paleosols often contain horizons rich in cylindrical peds that have been interpreted as burrows formed by cicada nymphs, whereas surface soils at the same sites lack burrows. Recent studies have shown that cicadas reached maximum numbers in association with a periglacial *Artemisia* (sagebrush) shrub steppe. The cicada-burrowed horizons, along with evidence from phytolith assemblages, suggest that these paleosols formed under cold, arid, full-glacial climates. Cicada-burrowed, late-glacial paleosols abruptly grade upward into worm-burrowed horizons of the modern surface soil associated with perennial grass-steppe vegetation and moister, warmer interglacial climates.

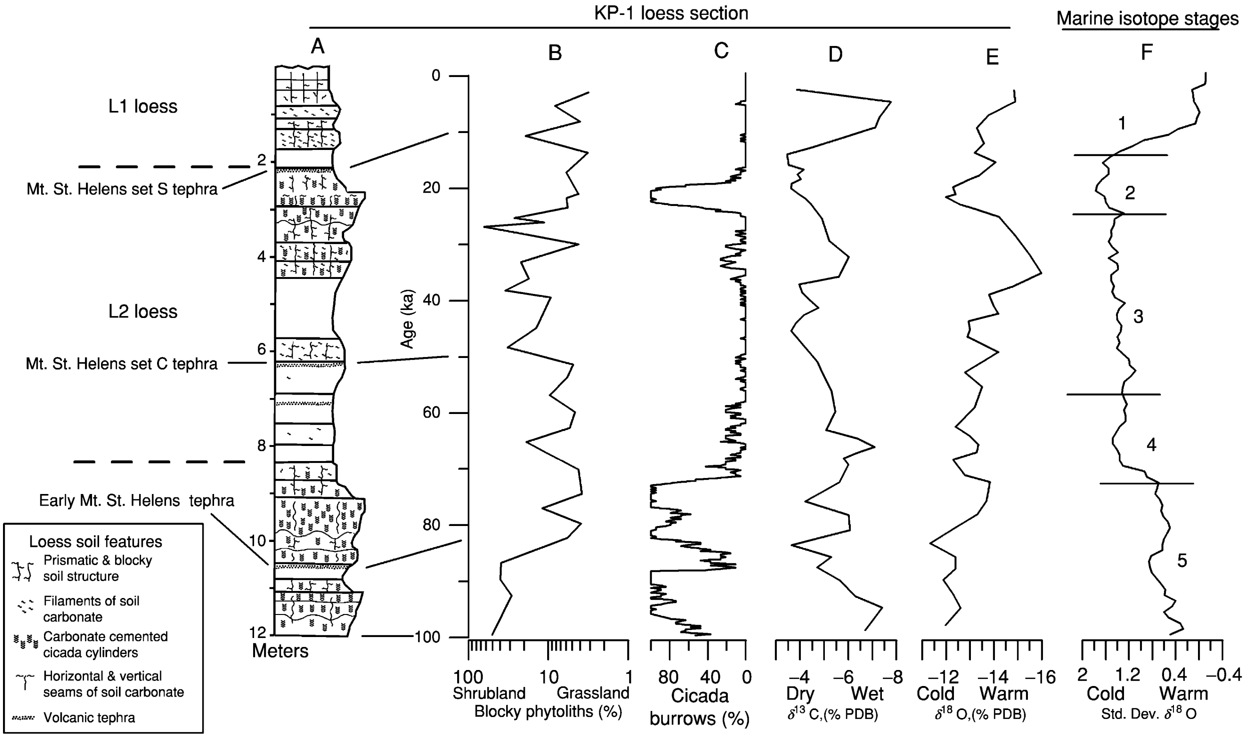


Figure 6 KP-1 loess (Palouse loess, US PNW) stratigraphy and paleoclimate proxies. A, loess stratigraphy and features of buried soils. Age–depth model of loess was derived from luminescence age dating. B, relative percentage of shrub (arid *Artemisia*) versus grass (moist perennial grasses) cover through time based on phytoliths preserved in the loess. D and E, change of carbon and oxygen isotopes (respectively) extracted from pedogenic carbonate. Palouse vegetation is approx. 100% C3 type. Carbon isotope signature is a proxy for paleowetness. Oxygen isotope signature is a proxy for paleotemperature. F, oxygen isotope curve obtained from marine sediments, separated into major isotope stages, with odd numbers representing interglacials and even numbers representing glacials. Cold and arid periglacial conditions in the Palouse indicated by relatively high abundance of shrubland phytoliths, cicada burrows, and positive isotopic signatures. Adapted with permission from Blinnikov M, Busacca A, and Whitlock C (2001) A new 100 000-year phytolith record from the Columbia Basin, Washington, USA. In: Meunier JD and Colin F (eds) *Phytoliths: Applications in Earth Sciences and Human History*, pp. 27–55. Lisse, The Netherlands: AA Balkema. © A.A.Balkema.

Carbon isotopes extracted from soil carbonate have been used as a proxy for temperature and moisture conditions. The isotopic signature can indicate the proportion of plants using the C_3 or C_4 metabolic pathway or, as in the example in Figure 6, signal changes in soil respiration rate in a C_3 plant community resulting from changes in precipitation. Oxygen isotopes also retrieved from soil carbonates and soil organic matter may provide information on paleotemperatures. Caution must be advised when interpreting carbon and oxygen isotope data from soil carbonates unless one can assess the possibility of atmospheric, groundwater, or other sources of contamination.

Magnetic susceptibility (MS) has been an important tool in correlating loess-paleosol sequences to deep-sea oxygen isotope records of paleoclimate. During soil formation, magnetic particles can concentrate via dust fall or due to the production of magnetite and maghemite via pedogenic processes, resulting in higher MS signals in paleosols than in unaltered loess. The depth function of MS in long loess-paleosol sequences in China in some cases matches strikingly well with paleoclimate records from deep-sea sediment cores.

Loess is generally too oxidizing for the preservation of plant matter, including pollen, but in some cases pollen analysis has been utilized. Phytoliths composed of opaline silica are formed in plant cells and are preserved in loess paleosols after the plants decay, providing a local, accurate assessment of past plant assemblages. Distinctive shapes can in some cases allow the identification of different genera of grasses, unlike pollen analysis. Phytolith analysis has been demonstrated as an important tool to reconstruct paleoclimate in the Palouse loess (Figure 6), New Zealand, and the Great Plains. Local assemblages can be compared with other local or regional proxy records of plant communities.

Surface Soils in Loess

European workers have repeatedly emphasized that loess is more than a sedimentary deposit; that its properties are a result of pedogenic weathering processes that impart porosity, unique structure, tawny color, and that reorganize calcium carbonates.

Large areas of loess soils such as the US Great Plains, Kazakstan, and the Ukraine, are in semiarid climates that supported grass-steppe vegetation prior to agricultural development. As a result, loess soils are dominated by the Mollisol soil order (US Soil Taxonomy System). (*See* **Civilization, Role of Soils**.) Where loess has accumulated in a forest setting or where hardwood or mixed forests succeeded on to loess areas after dust deposition ceased, gray moderately leached Alfisols with a clay-enriched argillic subsoil horizon are common. Soils formed in loess of true desert climates produce Aridisols with pale topsoil horizons and subsoil calcic horizons at shallow depth.

Historically it has been observed that loessial soils are very productive compared with soils formed in other parent materials. This is because silty soils provide high porosity and high water-holding capacity, which affords good aeration and high levels of plant-available water, relative ease of root penetration, and a stone-free seedbed. In comparison, sandier soils retain less moisture, and clayey soils retain more total moisture but have less plant-available water than loessial soils. Clayey soils pose additional problems, including limited depth of rooting and excessive wetness. High native fertility of loess soils is most notable in Mollisols, where nitrogen (N) is supplied by the mineralization of humus in the early years of cultivation; however, this must be supplemented by fertilizer N for continued high yields of cereal crops. Soil organic matter in loessial agricultural soils in the PNW US is less than half of native levels because of increased mineralization rates due to annual tillage and because of tillage erosion and water erosion. Loss of organic matter can lead to poor soil structure and increased soil crusting, which in turn can reduce seedling emergence and increase water erosion.

Environmental Concerns

Wind and water erosion have significantly degraded loessial soils worldwide. Erosion results in the loss of topsoil and soil nutrients, which have negative effects on productivity. Eroded soil materials can seriously degrade off-farm surface water, groundwater, and air quality. The North American Great Plains suffered severe wind erosion during the 'dust bowl' drought years of the mid-1930s that resulted in huge economic losses. In the Palouse area of the PNW US, wind erosion of agricultural fields has resulted in a 10-fold increase in atmospheric dust since the late 1800s. As of 2003, wind erosion from agricultural fields is a major producer of particulate matter less than 10 μm in size (PM_{10}), which poses significant heath risks to people of the region. Farmers have attempted to decrease wind erosion on fallow fields by leaving stubble, which increases surface roughness, and by the implementation of annual cropping. (*See* **Windbreaks and Shelterbelts**.) Loessial soils that are most prone to wind erosion are those with less than 1% soil organic matter content and less than 10% clay content.

Water erosion is a serious problem in loessial soils. Soil loss of up to 450 tonnes ha^{-1} per year has been recorded in autumn-planted winter wheat, where surface protection from rainsplash is minimal over winter. The erosion hazard can be increased by surface crusting and by frozen soils. Surface crusting occurs during rainfall, when raindrop impact detaches particles from soil aggregates and orients platy minerals such as micas. Crusting inhibits infiltration and encourages overland flow of water, which enhances erosion. Frozen soil similarly inhibits or prevents infiltration of rainwater. Snowmelt on frozen loessial soils may cause liquefaction of the thawed surface layer. (*See* **Crusts:** Structural.)

Unweathered loess has high porosity of 40–55% that decreases during pedogenesis due to translocation of particles, precipitation of minerals, and collapse of pores. Dry loess can form vertical walls several meters high, because it has open microfabric, cohesion that may be attributed to cementation of particles by carbonate, clay bridges between silt particles, preferential grain orientation, or attractive forces between small particles. Loading may result in the collapse of open pores, and frequent wetting may wash clay bridges from the silt particles, allowing failure of loess. Subsurface water flow results in piping, and the excavation of these pipes through time may result in collapse. Thick loess in humid regions is typically deeply gullied with high, steep walls. Tensile fractures may be common and can result in failure. Large landslides triggered by seismic activity have been documented in the thick Chinese loess.

Dust Additions to Other Soils

In places where dust accumulation rates are not large enough to result in loess deposits, dust additions still may be of great importance to soil fertility. Dust additions to soils were originally recognized by the presence of minerals such as quartz in places where the bedrock such as basalt contained none. Terra Rossa soils on limestones have been argued to form by very slow dust accretion over eons rather than by residual enrichment of noncarbonate minerals in the limestone. Large dust plumes from Asia and Africa have been well documented to travel in the upper atmosphere around the globe and deposit thousands of kilometers from their source areas. Far-traveled Asian dust is a major contributor to soils in Japan, Korea, and Hawaii and has traveled as far as North America. The same can be said of Saharan dust, which contributes to soils of the Canary Islands and southern Europe. Saharan dust plumes also make their way to South and North America.

Biogeochemical studies of ecosystem development have shown that, in areas of highly weathered soils, productivity is sustained by the deposition of exotic mineral dust, which can supply critically limiting nutrients such as phosphorus and potassium. Rainforest soils of the Amazon in South America include highly weathered Ultisols and Oxisols. These old soils can no longer provide nutrients from weathering of parent material, yet they sustain a diverse ecosystem. It has been shown that African dust is a significant source of nutrients to the Amazon Basin. A similar situation exists for old, weathered tropical soils of Hawaii, which receive an influx of nutrients from Asian dust. Dust has been shown to be a major source of N and P and organic matter in soils of Australia and Israel. Soluble salts in dust may have an impact on soil pH.

On a geologic time scale, additions of calcium carbonate and silica from dust greatly aid the formation of duripans and petrocalcic horizons (cemented horizons of silica and calcium carbonate, respectively). (*See* **Dryland Farming; Productivity; Remote Sensing:** Organic Matter; **Sand Dunes**.)

List of Technical Nomenclature

Alfisol	A soil order in the US system of soil classification; a mineral soil, generally forested, that has a light-colored surface horizon and clay-enriched subsurface horizon that is not highly depleted in plant nutrients
Argillic horizon	A subsurface mineral soil horizon characterized by the illuvial accumulation of clays; a subsurface horizon of clay enrichment by the downward movement of clays
Aridisol	A soil order of the US system of soil classification; a mineral soil that formed in a desertic climate and that has some development of soil horizons, often a calcic or duripan horizon
Biogeochemical	In reference to systems influenced by a combination of biological, chemical, and geological processes
Calcic horizon	A subsurface mineral soil horizon characterized by enrichment in calcium carbonate, often by the downward movement of ions with precipitation of the carbonate in a lower horizon
Cambic horizon	A subsurface mineral soil horizon characterized by the alteration or removal of mineral materials as indicated by mottling or gray colors, or by more intense or redder colors than the parent material from which the soil formed, or by the removal of carbonates

Duripan	A subsurface mineral soil horizon cemented by illuvial silica; a type of hardpan
Eolian	In reference to entrainment, transport, and deposition of sedimentary particles by the wind
Greda	A European term for a linear topographic ridge in a loess-covered landscape
Holocene	An epoch of the Quaternary period beginning at the end of the ice ages approximately 10 000 years ago and extending to the present day
Horizon	A layer of soil, usually parallel to the land surface, with a distinctive set of physical, chemical, and biological properties and differentiable from horizons above or below it by differences in appearance and properties
Illuviation	A soil-forming process involving the movement of soluble or particulate soil materials from one horizon downward into another horizon (cf. argillic horizon, calcic horizon)
Luminescence dating	A radiometric age-dating technique that measures the time elapsed since a mineral grain was last exposed to sunlight. Most often applied to age-dating a sample of eolian sediment
Magnetic susceptibility	The ratio of induced magnetization to the strength of the magnetic field causing the magnetization
Micromorphology	The study of undisturbed soil samples by light microscopy from specially prepared microscope slides, called thin sections
Mollisol	A soil order of the US system of soil classification; a mineral soil that has a dark-colored humus-rich surface horizon and one of several kinds of subsurface horizons
Oxisol	A soil order of the US system of soil classification; a mineral soil that has an oxic horizon, which is highly weathered, infertile, and is highly enriched in kaolin clays or in iron and aluminum oxide minerals
Paleoclimate	Referring to climate conditions of times prior to today
Paleosol	A buried soil profile formed in an environment of the past or a surface soil that has undergone weathering during several climatic episodes
Pedogenesis	The physical, chemical, and biological processes of soil formation
Petrocalcic horizon	A subsurface mineral soil horizon characterized by cementation by calcium carbonate
Phytolith	Microscopic structures of opaline silica formed within the tissues of living plant. Phytoliths extracted from soils or sediments are sometimes used like fossil pollen to infer vegetation or climates of the past
Pleistocene	An epoch of the Quaternary period that began approximately 2.4 million years ago and ended approximately 12 000 years ago. A time period corresponding to the ice ages
Porosity	The ratio of the volume of void or pore space to the total volume of a rock or sediment
Quaternary	A period within the Cenozoic era consisting of the Pleisocene and Holocene epochs
Saltation	The transport of sand-sized sediment particles in which grains are moved by wind in a series of short leaps or bounces
Stable isotopes	One of two or more species of the same nonradiogenic chemical element having the same number of protons but differing in the number of neutrons. The isotopes of an element have slightly different physical and chemical properties owing to their mass differences
Strata	Layers of sedimentary rock
Suspension	The transport of generally silt-sized and finer sediment particles in air whose velocity or turbulence allows the particle to remain for a time aloft
Tephra	Any rock materials ejected from a volcano and transported through the air, including pumice and ash
Ultisol	A soil order of the US system of soil classification; a mineral soil that has a clay-enriched subsurface horizon that is highly depleted in plant nutrients
Welded soil	Two or more soils of different ages in a depositional sedimentary system that are superimposed on one another and partially overlap

See also: **Civilization, Role of Soils**; **Dryland Farming**; **Erosion:** Wind-Induced; **Minerals, Primary**;

Morphology; **Productivity**; **Remote Sensing:** Organic Matter; **Sand Dunes**; **Windbreaks and Shelterbelts**

Further Reading

Blinnikov M, Busacca A, and Whitlock C (2001) A new 100 000-year phytolith record from the Columbia Basin, Washington, USA. In: Meunier JD and Colin F (eds) *Phytoliths: Applications in Earth Sciences and Human History*, pp. 27–55. Lisse, The Netherlands: AA Balkema.

Busacca AJ, Beget J, Muhs DR *et al.* (2003) Eolian sediments. In: Gillespie AR, Porter SC, and Atwater BF (eds) *The Quaternary Period in the United States. Developments in Quaternary Science 1*, pp. 275–310. Amsterdam: Elsevier.

Catt JA (2001) The agricultural importance of loess. *Earth Science Reviews* 54: 213–229.

Chadwick OA, Derry LA, Vitousek PM, Huebert BJ, and Hedin LO (1999) Changing sources of nutrients during four million years of ecosystem development. *Nature* 397: 491–497.

Frazee CJ, Fehrenbacher JB, and Krumbein WC (1970) Loess distribution from a source. *Soil Science Society of America Proceedings* 34: 296–301.

Kemp RA, Derbyshire E, Xingmin M, Fahu C, and Baotian P (1995) Pedosedimentary reconstruction of a thick loess-paleosol sequence near Lanzhou in north-central China. *Quaternary Research* 43: 30–45.

Livingstone I and Warren A (1996) *Aeolian Geomorphology*. Singapore: Addison Wesley Longman Ltd.

Muhs D and Zárate M (2001) Late Quaternary eolian records of the Americas and their paleoclimatic significance. In: Markgraf V (ed.) *Interhemispheric Climatic Linkages*, pp. 183–216. London: Academic Press.

Pye K (1987) *Aeolian Dust and Dust Deposits*. London: Academic Press.

Pye K and Sherwin D (1999) Loess. In: Goudie AS, Livingstone I, and Stokes S (eds) *Aeolian Environments, Sediments and Landforms*, pp. 213–238. Chichester, UK: Wiley & Sons.

Wright JS (2001) 'Desert' loess versus 'glacial' loess: quartz silt formation, source areas and sediment pathways in the formation of loess deposits. *Geomorphology* 36: 231–256.

LOWDERMILK, WALTER CLAY

J D Helms, USDA Natural Resources Conservation Service, Washington, DC, USA

Introduction

Walter Clay Lowdermilk (Figure 1) often described his profession as "reading the records which farmers, nations, and civilizations have written in the land." While others have belonged to this profession, certainly few had the inclination, ability, and opportunity to indulge in it, as did Lowdermilk. The profession required expertise in many fields of study, but as practiced by Lowdermilk it was not a purely academic exercise. Rather, he sought an ambitious objective – a permanent agriculture for the world. He had begun his career researching the basic processes of erosion and runoff, especially as influenced by humans. Lowdermilk became a member of the early-twentieth-century conservation movement in the USA, a movement with a strong scientific bent. Its scientists held that treatment of natural resources should be in accordance with scientific principles, not propelled by emotionalism or untested theories. A leader in organizing government conservation agencies in the USA and abroad, Lowdermilk put research findings into action. He enunciated plans for agricultural development in *Palestine, Land of Promise*. Later, he worked in Israel to write the plans on to the face of the countryside. As an author he spoke for the conservation

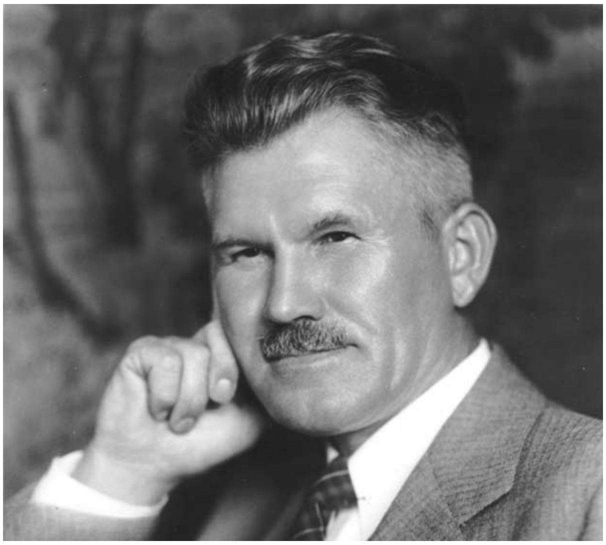


Figure 1 Walter Clay Lowdermilk.

movement, and in his international work served as a humanitarian by furthering conservation and the proper use of land to make and keep it productive.

Lowdermilk, Walter Clay (1 July 1888 to 6 May 1974), soil scientist, geologist, soil conservation leader, and author, was born in Liberty, North Carolina, the son of Henry Clay Lowdermilk, a businessman, lumberman, and rancher, and Helen Vashti Lawrence Lowdermilk. The family moved westward to Missouri, to Oklahoma, and finally to Arizona. Walter Lowdermilk graduated from the Park College Academy in Parkville, Missouri, in 1906 and then attended Park College (1908–10). In 1910 he enrolled at the University of Arizona; after 2 years there he won a Rhodes Scholarship to Oxford University, where he earned a BS degree in forestry (1914); a BA degree in geology (1915); and an MA degree, granted *in absentia* (1922). While at Oxford he had an opportunity to study forestry in Germany. He also served on Herbert Hoover's Commission for Relief in Belgium (1914–15). After the scholarship years, he returned to the USA as a ranger in the Southwest for the Forest Service. During World War I, he returned to Europe as timber acquisitions officer and then measured timber for assessing war-time damages. Returning from World War I, he became the Forest Service's regional research officer in Montana.

A man who enjoyed research work, he had found a position that offered satisfaction. Given his ability, there was opportunity for advancement. But he was not to remain on that career ladder. Soon he would be in China, where, he later recalled, the significance of soil erosion was burned into his consciousness. Through the years in England and afterward, the young forester had corresponded with Miss Inez Marks, a friend from Arizona. On leave from her missionary work with the Methodist Church in China, she agreed to meet him at the Rose Bowl, New Year's Day, 1922. Marriage plans quickly followed. Her entreaties that China desperately needed talented scientists led to his applying for a position with the University of Nanking's school of agriculture and forestry. The couple married in August and departed for China in September 1922. Lowdermilk's charge, for a small salary, was to assist in solving the flooding problems and resulting famines. Exactly how a forester was to help with food production remained a mystery as he attended university classes to learn Mandarin during the first year.

An expedition to the Yellow River solved the mystery. There he stood atop a section of the 650-km-long dike that held the river 12–15 m above the flood plain. This marvel was a result of Chinese labor necessitated by silting of the river's channel – 'aggradation' in the terms of earth scientists. Lowdermilk set out to find the source of the silt. In spring 1924, O.J. Todd, engineer of the International Famine Relief Commission, accompanied Lowdermilk on a 3000-km trip to the watersheds of the Yellow and Wei rivers. Todd's mission was to study the Wei-Peh irrigation project. Few foreigners had visited northwest China, where the pair completed a third of the journey afoot or on mulecart or muleback. In Shensi province, they found a plateau consisting of deep, undulated deposits of loessial soils. Depth, fertility, and erodibility made these fine, wind-deposited soils prime locations for man-induced erosion. In the deforesting activities of the people, Lowdermilk found the reason for the gigantic 200-m-deep gullies. The Chinese shaved the hills clean of all herbaceous shrub and tree growth for firewood. Paradoxes abounded on the trip. Temple forests, reproduced naturally and protected by Buddhist priests, provided evidence of the denuded hills' capability for sustaining vegetation. Bench terraces festooned some slopes. Yet some of the best agricultural land on the level, alluvial plains was used for timber production under irrigation. Surrounding hills were little used for timber.

The pair visited Sianfu, the capital city of China during its Golden Age, where Todd wanted to inspect the irrigation works. The area retained little of its former prosperity, which Lowdermilk conjectured had flowed from a great irrigation project, long since silted up. The forester returned to his post at the University of Nanking with an impression of colossal erosion contrasted with evidences of former grandeur. Already he had decided to expand his study of the sciences involved with natural resources to include the actions of people as well as man's relationship to nature.

Historical research revealed that the Yellow River had changed course eight times since AD 11. Several times the river had been restrained by dikes, only to break free. Once it emerged 600 km from its former outlet. Dikes, therefore, were essential to using the plain for agriculture. But building higher dikes, Lowdermilk concluded, was not a lasting solution unless the aggradation of the river was reduced by checking the supply of silt. Lowdermilk's supposition that erosion caused frequent and severe flooding had been recognized in the USA, but only on the small water courses, not on the lower reaches of major rivers. The China experience – siltation of a major river channel as a cause of flooding and channel relocation – was on a scale unknown in the USA.

Lowdermilk's recommendation for flood control gave some indication of the breadth of his training in sciences, especially geology, and his ability to assimilate the findings into a solution. The Yellow River and its tributaries had excavated a deep channel into the plateau created by the wind-deposited soils. Recognizing that removal of vegetation allowed runoff to carve gullies in the loessial plain and that gully wash accounted for most of the silt, he proposed attacking erosion by planting trees on the talus slopes at the foot of the gullies. The forested gullies would be guarded and managed by villages as community forests to provide wood. Undissected portions of the loessial plateau could be used for agriculture. Where and when possible, check dams should be used to raise the base level of streams and prevent incision by the gullies farther into the plateau. Treatment of the watershed was directly tied to famine prevention. He concluded that soil and water conservation were urgently necessary to increase the productivity of this region of China.

Lowdermilk was not content to base his recommendations exclusively on empirical evidence. Certainly the scientific forestry school, whence he came, demanded another explanation. Using the runoff and erosion plot method devised by F.L. Duley and M.F. Miller at the University of Missouri, he and his Chinese associates set up plots on 20 temple forests and on denuded areas for comparison. After 3 years of study, he presented the findings. Runoff from denuded areas greatly exceeded that of temple forests or areas reclaimed through reforestation. The main reason for the excess runoff, he believed, was that particles of soil on the denuded areas clogged the pores of the soil surface. Forest litter arrested this action.

Further study convinced Lowdermilk that 40–60% of the uplands in northern China had little cover to retain runoff. So great had been the rapid runoff that it had reduced evaporation and brought on a period of decreased precipitation in the area. With this argument, Lowdermilk projected a hypothesis that he would later apply to other lands. Scholars had long been presented with anomalies of twentieth-century poverty contrasted with evidences of former civilizations which possessed a high degree of culture and prosperity. Some scholars, notably Ellsworth Huntington and Baron Von Richthofen, found the answer in climatic change. In the case of north China, Lowdermilk not only saw soil erosion and flooding as the reason for decline, but also claimed their effects as the reason for a climatic change.

The communist uprising of 24 March 1927 in Nanking ended the Lowdermilks' stay in China. Leaving behind all possessions, they barely escaped. At the University of California, Lowdermilk combined study for a PhD from the School of Forestry (minors in soil science and geology) with research at the California Forest Experiment Station. Here he reentered the fray over the effects of vegetative cover on runoff, erosion, and flooding. On one of his treks in China, Lowdermilk had heard the proverb "Mountains empty – rivers gorged." He judged the application of timber management in that locale to be superior to any system he had observed in Germany. The Chinese and other civilizations had recognized the value of forest cover and acted upon their observations. Scientists in the conservation movement, however, demanded more than proverbs for proof, and the influence of forest cover on soil erosion and streamflow continued to be debated by hydrologists, engineers, and foresters.

In his study of the influence of forest litter on runoff and erosion, he used rain-making machines, soil profiles transferred to tanks, and measuring instruments of his design. In 1929, he presented the confirmation for what he and others had observed. On bared soil the raindrops splashed up muddy. As muddy water percolated into the soil profiles, suspended particles were filtered out at the soil surface. The thin layer thus formed reduced percolation and increased runoff. The water-absorbing capacity of forest litter had little influence on runoff. However, the litter filtered the water and kept the soil profile open to percolation. The experiments confirmed a hypothesis that Lowdermilk had first presented at the Third Pan-Pacific Science Congress in 1926 in Tokyo. He also proposed terms for classifying erosion. 'Accelerated erosion' arose from the 'artificial disturbance of factors which controlled the development of soil profiles.' In the absence of such disturbances, one could view any erosion as the 'geologic norm of erosion.'

As Lowdermilk was completing his dissertation, another research opportunity developed with the Forest Service in California. Partly in response to the prodding of Hugh Hammond Bennett, a US Department of Agriculture (USDA) soil scientist, Congress, in the Agricultural Appropriations Act of 1930, authorized a group of experiment stations to study erosion and runoff. Bennett supervised most of the stations from his post in the Bureau of Chemistry and Soils, but the Forest Service also received funds for stations. Lowdermilk designed the research station and program on a large, isolated watershed in southern California. The San Dimas watershed provided an excellent opportunity to test the effects of watershed management on water yield. Expanding towns and citrus orchardists at the foot of the watershed had to dig increasingly deeper wells to reach underground aquifers. Whether the vegetative mantle should be burned to reduce transpiration or protected from fire for maximum ground water supplies was a matter of controversy. To demonstrate and measure the relationship of percolation to aquifer levels, Lowdermilk had Civilian Conservation Corps (CCC) enrollees build water-spreading structures, which led to a gravelly basin where the silt settled out and water percolated to the aquifers.

Experiments at San Dimas also contributed to a better understanding of soil genesis processes such as horizonization and the anthropogenic influence on soils in a densely populated region. CCC enrollees sifted and mixed soil, which was then placed in lysimeters. They also kept soil samples in glass containers. Soil scientist Robert C. Graham and colleagues examined the uncontaminated, glass-enclosed samples and compared them with soils in the lysimeters. Lowdermilk had read and cited the major soil scientist treatises of the day by Harry O. Buckman, Jacob S. Joffe, Jacob Lipman, Curtis F. Marbut, G.W. Robinson, and Nathaniel Southgate Shaler; but he did not specialize in topics such as soil genesis, soil mineralogy, or soil chemistry, all topics which were being pursued in the new soil science departments at land-grant universities. Rather his studies addressed the factors that prevented erosion and other soil degradation, and thus preserved the upper soil environment for agriculture.

While working on the San Dimas project, Lowdermilk was recruited to be the vice-director of a temporary New Deal agency in the Department of the Interior, the Soil Erosion Service, in October 1933. The aforementioned Hugh Hammond Bennett had been selected as Director. The news service commenced the campaign for adopting conservation practices on American farms through demonstration projects based in small watersheds. Young soil conservationists worked with farmers to develop conservation plans for the farms. The plans had to achieve both the conservation objectives as well as satisfy the farmers' needs. Enrollees in CCC camps located on the projects and Works Progress Administration workers supplied labor. Later the service expanded operations nationwide by assigning trained conservationists to work with locally organized conservation districts. An act of 27 April 1935 created the Soil Conservation Service (SCS) in the USDA and gave it continuing authorities and duties. Lowdermilk served as associate chief (Dec. 1935 to Mar. 1937), chief of the research division (Mar. 1937 to Nov. 1939), and then head soil conservationist until his retirement.

Lowdermilk contributed to organizing all facets of the new service, but he especially concentrated on the research program. USDA transferred the experiment stations established under the 1930 Agricultural Appropriations Act to SCS in 1935. The stations were already engaged in research on terracing, crop rotations, strip-cropping, tillage methods, and the value of all these to soil conservation. Lowdermilk added runoff and erosion studies. Furthermore, the demonstration projects afforded an opportunity to observe the effects of farmland conservation practices on the entire watershed. Lowdermilk was thus instrumental in the new agency's contributions to understanding small-watershed hydrology.

In 1938 chance again intervened in Lowdermilk's life; and as usual he seized the opportunity. Representative Clarence Cannon suggested that a survey of the Old World could be useful in the United States' efforts toward a permanent agriculture. The trip, August 1938 to November 1939, involved more than 25 000 miles of automobile travel in Europe, the Mediterranean area, and the Middle East. Here he perfected his art of reading the land for evidence of past use and misuse. Before undertaking surveys in each country, Lowdermilk consulted agriculturalists, scientists, and officials. Geologists and archaeologists were especially interested, and valuable to Lowdermilk in explaining the cultural and physical factors involved in land use. In addition to searching for soil conservation and flood prevention measures that might be imported to the USA, Lowdermilk was engaged in what he called 'agricultural archaeology.' Ruins of some preindustrial civilizations indicated a prosperous agriculture, although these areas now had serious resource problems. What events brought about such conditions? What were the lessons for contemporary civilizations?

Lowdermilk's land-read records of past civilizations appeared in numerous articles. Indeed, there were 'Lessons from the Old World to the Americas in land use,' as Lowdermilk titled an article in the annual report of the Smithsonian Institution. He gladly noted the cases of wise land use through centuries, but was usually obliged to find a story of deterioration. The Soil Conservation Service published a summary, *Conquest of the Land Through 7000 Years*, in 1953 and followed it with several reprintings until more than one million copies were distributed. Early on, Lowdermilk had learned the value of written communication and practiced his writing skills. Readers who know Lowdermilk only through this publication have perhaps a truncated view – that of the globe-trotting chronicler of calamities awaiting civilizations that abuse their resources. He realized that a civilization's decline could not be interpreted solely on the basis of soil erosion. However, in writing the pamphlet, he embarked on a didactic mission aimed at all Americans, not just farmers. Soil fertility was a matter of concern for the farmer. Maintaining the medium for fertility – the physical body of soil resources – concerned the entire nation. Without it, 'liberty of choice and action' was gone.

World War II terminated the trip in Europe but it opened a new opportunity, a return to China. At the behest of the Chinese government, Lowdermilk undertook the dangerous journey to advise the Chinese about increasing their food supply. The Communist triumph in China meant the Lowdermilk's and other advisors' plans for postwar-China would not be acted upon. But the trip afforded Lowdermilk an opportunity to pursue his interest in agricultural archaeology. During the intervening years in the USA, he had continued to study the agricultural archaeology of China. While in China he bought gazetteers, local histories, which Dean R. Wickes, a Chinese-language specialist, then researched for evidences of erosion problems. This research showed that in northern China, an area with a small percentage of level land, the population had increased threefold since the mid eighteenth century. This rapid population increase sent people to the hills for firewood and arable land, without any orderly installation of engineering measures for soil conservation. Unlike areas of central and southern China, they had no elaborate bench terraces to protect farmland. The gazetteers provided accounts of clearing the slopes, removing farmland from the tax rolls as wasteland, and abandoning homes along streams due to frequent flooding.

Lowdermilk retired from the SCS on 30 June 1947, but he continued his international work, which had been fairly continuous since the 1938 trip to the Middle East. He consulted in French North Africa (1948) and in the British colonies in Africa (1949–50). In the USA he contributed to the President's Water Resources Policy Commission. Controlling erosion on the upper reaches of watersheds became a passion for Lowdermilk's generation of conservationists. They favored land cover for increased absorption and engineering works for the controlled disposal of water without erosion. The upstream reservoir on the small watersheds was an integral part of the river development – an assertion that was often contested. Proponents of the control and use of headwaters had stated their case in the publication *Little Waters and Headwaters: Control and Use*. In the late 1940s, conservationists had another opportunity when Morris Cooke, a force behind *Little Waters*, became chairman of the President's Water Resources Policy Commission. Lowdermilk assumed chairmanship of the Committee on Standards for Basic Data. The Cooke and Lowdermilk views held sway in the committee report that emphasized a comprehensive, interdisciplinary approach. The interdependence of land and water called for watershed management, which had been neglected due to "our natural endowment and relatively low population density." Furthermore, the small watershed, the unit of watershed management preferred by the authors, was a cultural unit. The watershed unit had to be small enough so that residents understood its influence on their lives. Then they would devote the time and money needed to bring it to fruition as a community watershed. Lowdermilk's experience in semiarid climates came through in the committee's attitude toward flood control. Where feasible, reservoirs should not be used solely to control floods, but also to store storm waters for later use. Following work with the US Water Resources Policy Commission (1950) and the Supreme Allied Command in Japan (1951), he worked in Israel under the auspices of the Food and Agriculture Organization of the United Nations to organize an SCS.

During the trip to the Middle East in 1938–39, Lowdermilk became inspired by the efforts of urban-born European Jews to reclaim land. Upon returning to the USA, he wrote *Palestine, Land of Promise* (1944), wherein he proposed a Jordan Valley Authority to provide power, water supply, and irrigation water. The book argued, as did his popular pamphlet *Palestine Can Take Millions* (1944), that the land there could support a much larger population. The message contravened the British and American foreign policy of the early 1940s that discouraged Jewish emigration to Palestine. After retirement from the SCS, he worked with the Israelis to implement some of the measures outlined in the book. Drawing upon his Western experience, he advocated diversion of the Jordan River water before it entered the Dead Sea. The water would be utilized for hydroelectric power and irrigation water. Many Israelis favored technical assistance for agricultural development over direct food assistance. That sentiment was concisely conveyed when Minister of Development, Mordecai Bentov, coined the saying, "We don't need powdered milk; we need Lowdermilk." While there, Lowdermilk helped establish at Haifa a school to train conservationists, a school which later bore his name. The Lowdermilk School of Agricultural Engineering emphasized the basic sciences as preparatory to agricultural studies. Students took 2 years of mathematics, chemistry, physics, geology, and biology before moving on to the agricultural sciences. A job-related project in the 5th year was necessary to earn the degree.

Lowdermilk became best known for his work in Israel, probably more so than even for his work in China, at San Dimas, or with the SCS. In a career that has generally been lauded, recommendations for development in the future state of Israel occasioned the most criticism of his work. A progressive conservationist who came of age in the early twentieth century, he advocated what he saw as positive human intervention to correct past detrimental human actions. In the Middle East he usually discounted climatic causality for land degradation and emphasized the human element. He singled out the nomadic herders as a cause for soil erosion. In place of nomadic herding, he would substitute development of the Jordan River to supply electric power and irrigation water for intensive agriculture and regional economic development. The new state of Israel needed such development to support the immigrants, he argued. In the USA he and Mrs. Lowdermilk spoke for Israeli statehood and were counted as 'Christian Zionists.' Lowdermilk cited an area he knew well, southern California, as an analogy of economic development brought on by the combination of hydroelectric power and irrigation water. While he was careful to say that Arabs would also benefit from economic development, none the less such arguments could be seen as antithetical to Arab interests. During recent decades environmentalism and the study of environmental history have generally looked with some disfavor on the progressive conservationists' merging of conservation with economic development. Some favor indigenous, less-intensive systems such as nomadic grazing as more appropriate.

Lowdermilk received many professional honors. He was a fellow and president of the American Geophysical Union (1941–44) and a fellow of the Society of American Foresters and of the Soil Conservation Society of America. The Lowdermilks had two children, William Francis and Winifred Esther. In retirement the Lowdermilks had a home in the Berkeley hills, California. Walter Lowdermilk died in Palo Alto, California in 1974, having spent his last years at the Veterans Hospital there.

See also: **Archeology in Relation to Soils**; **Climate Change Impacts**; **Degradation**; **Desertification**; **Erosion:** Water-Induced; **Jenny, Hans**; **Loess**; **Marbut, Curtis Fletcher**; **Sustainable Soil and Land Management**; **Terraces and Terracing**; **Water Harvesting**; **Watershed Management**

Further Reading

Chall M (ed.) (1969) *Walter Clay Lowdermilk, Soil, Forest, And Water Conservation and Reclamation in China, Israel, Africa, and the United States*. Oral History Interview. Bancroft Library, University of California–Berkeley, CA.

Davis RCB (1999) *Development as a Tool of Diplomacy: The Domestic Models for United States Policy in the Jordan River Valley, 1939–1956*. PhD dissertation. Georgetown University, Washington, DC.

Egerton-Warburton LM, Graham RC, Allen EB, and Allen MF (2001) Reconstruction of historical changes in mycorrhizal fungal communities under anthropogenic nitrogen deposition. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 268: 2479–2484.

Graham RC and Wood HB (1991) Morphologic development and clay redistribution in lysimeter soils under chaparral and pine. *Soil Science Society of America Journal* 55: 1638–1646.

Helms JD (1984) Walter Lowdermilk's journey: forester to land conservationist. *Environmental Review* 8: 132–145.

Lowdermilk IM (1985) *All in a Lifetime: An Autobiography*. Berkeley, CA: Lowdermilk Trust.

Lowdermilk WC (1944a) *Lessons From the Old World to the Americas in Land Use*. Annual Report of the Board of Regents of the Smithsonian Institution, 1943, pp. 413–427. Washington, DC: Government Printing Office.

Lowdermilk WC (1944b) *Palestine, Land of Promise*. London, UK: Victor Gollancz.

Lowdermilk WC (1953) *Conquest of the Land Through 7000 Years*. Agriculture Information Bulletin No. 99. Washington, DC: US Department of Agriculture.

Lowdermilk WC and Wickes DR (1942) Ancient irrigation brought up to date. *Scientific Monthly* 55: 209–225.

Lowdermilk WC and Wickes DR (1943) China and America against soil erosion. *Scientific Monthly* 56: 393–413; 505–520.

Stross RE (1986) *The Stubborn Earth: American Agriculturalists on Chinese Soil, 1898–1937*. Berkeley, CA: University of California.

LYSIMETRY

T A Howell, USDA Agricultural Research Service, Bushland, TX, USA

Published by Elsevier Ltd.

Introduction

Lysimeters are standard instruments used in hydrology and water-quality research. In particular, they are used to measure percolation beneath the vegetation root zone and water use through evaporative processes from vegetation. (*See* **Evaporation of Water from Bare Soil**; **Evapotranspiration**.) They can also measure 'net' additions to the soil from precipitation or irrigation.When used to measure percolation, they become *in situ* water-quality sampling instruments. (*See* **Drainage, Surface and Subsurface**; **Leaching Processes**; **Macropores and Macropore Flow, Kinematic Wave Approach**; **Salination Processes**; **Salinity:** Management; **Solute Transport**; **Vadose Zone:** Hydrologic Processes.)

The word 'lysimeter' is derived from the Greek roots 'lysis,' which means dissolution or movement, and 'metron,' which means to measure. The standard definition of a lysimeter is: a device for measuring the percolation of water through soils and for determining the soluble constituents removed in drainage; so clearly the *lysimeter* is a soil physics and hydrologic measurement tool. It defines the water flux at some arbitrary boundary beneath the soil surface so that water balance can be measured or a water sample extracted that represents the naturally percolated chemical constituents. As a water-balance measurement device, lysimeters are used to determine evaporation from the soil–plant system, commonly called evapotranspiration (ET), and/or the latent heat flux in a surface energy balance.

History

Lysimeters are reported to have been used in the Netherlands in the early seventeenth century and in France later in the seventeenth century. Most of these reported studies were focused on measuring crop water use. Lysimetry for determining water use from vegetation has advanced through better knowledge of soil physics of water flow, soil hydraulic and thermal processes, plant root water uptake, and micrometeorologic influences from advection and aerodynamic processes (fetch related).

Lysimeters can be used in a laboratory, greenhouse, environment chamber, or situated in the natural outdoor environment. When used to simulate natural outdoor environments, design details and site selection are critical, together with management, so the vegetation within the lysimeter duplicates the surrounding vegetation. In practical terms, a well-managed lysimeter is indistinguishable from its surroundings. When used in greenhouse, laboratories, or environmental chambers, lysimeters are more 'relative' than 'absolute' measurement devices. The physical and chemical characteristics of the lysimeter soil should duplicate the natural states under study.

Construction materials and drainage methods are critical in lysimetry used to determine soil-water chemical transport or solute fluxes. Care is exercised to select design materials that will not react with or absorb any of the chemicals that are of interest.

There is not a standard for lysimeters that can be followed to construct either a lysimeter for measuring chemical transport or for measuring water evaporation from soils or vegetation. However, one principle has emerged to guide design considerations, which is that the lysimeter must be representative of the conditions desired to be measured. If the lysimeter design, construction, and its operation cannot permit the vegetation or soil condition within the lysimeter to duplicate or accurately represent the environment under study, then reliable measurements of neither the evaporative processes nor the soil-water chemical transport processes can be expected.

Lysimeter Types

Lysimeters are foremost devices, typically tanks or containers, that define a specific boundary to contain soil water and permit measurement of either the soil-water balance or the volume of water percolating vertically and/or its quality. Lysimeters can be broadly defined as percolation lysimeters (sometimes called drainage lysimeters) or weighing lysimeters (Figure 1). Both can serve the purposes of determining the soil-water balance, the vertical percolation flux

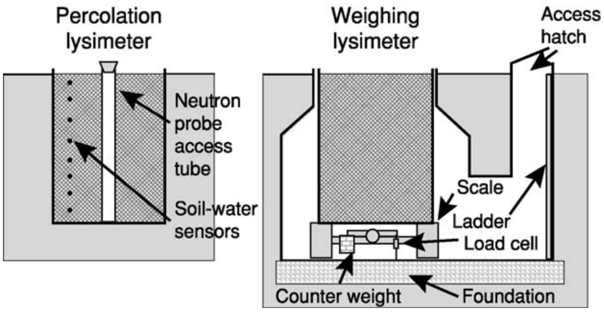


Figure 1 Schematic drawings of a percolation (drainage) lysimeter and a weighing lysimeter (not to scale).

(drainage), and the chemistry of the percolating water.

Percolation lysimeters may be installed to determine simply the vertical soil-water flux (drainage) or the chemical movement within the soil at a defined boundary. Soil-water solution sampling lysimeters can be simple porous devices that can extract soil water under a vacuum or access wells to permit water sampling from beneath a free water surface (i.e., water table). Percolation and weighing lysimeters differ in their measurement methods to determine the vegetation water use and/or soil-water evaporation. Percolation lysimeters must be used with a soil-water profile measurement method (*See* **Neutron Scattering**; **Tensiometry**; **Time-Domain Reflectometry**) to estimate 'indirectly' the water use in evaporative processes. Weighing lysimeters permit the mass or volumetric soil-water content change to be determined by weighing the lysimeter and determining its mass change over time. Weighing lysimeters, thus, can determine the 'net' infiltration from rainfall or irrigation and the amount of evaporation between wetting events. If properly designed, weighing lysimeters can coincidently measure the drainage rate as well as the evaporation rate. Percolation-lysimeter accuracy of the evaporative water balance is directly related to the precision of the soil-water measurement and its integration through the vegetation root zone. Weighing-lysimeter precision depends on many factors, the main ones being scale resolution, counterbalancing, and area-to-volume ratio.

Lysimeter Water Balance

The water balance of a lysimeter is given as:

$$\theta_{\mathrm{i}} = \theta_{\mathrm{i-1}} + P_{\mathrm{i_i}} + I_{\mathrm{i}} - D_{\mathrm{i}} - \mathrm{ET}_{\mathrm{i}} \qquad [1]$$

where: θ_{i} is the lysimeter water volume per unit lysimeter area, in millimeters, at time i; P_{i} is precipitation, in millimeters; in I_{i} is irrigation, in millimeters; D_{i} is drainage from the lysimeter, in millimeters; and ET_{i} is evaporative loss (evapotranspiration) from the lysimeter, in millimeters, all during the time interval $i-1$ to i. Typically, lysimeters have a rim height of 10–40 mm above the soil surface to prevent runoff and runon. Some lysimeters used in hydrology are equipped with tanks to capture volumetrically runoff above a specified depth of surface storage. If runoff does not occur from the lysimeter, the lysimeter 'freeboard' may increase rainfall or irrigation capture, thereby increasing water available for evaporation or adding to lysimeter percolation volumes. Generally, P and I are measured separately using various methods (i.e., rain gauges, flowmeters, etc.), while D can be measured volumetrically as drainage occurs from the lysimeter.

Although water can move upward, most lysimeters are not equipped to measure or add water to simulate upward water fluxes. Some lysimeters use a static or dynamic water table to mimic field water conditions, and these lysimeters are usually equipped to measure the water additions needed to maintain the water table depth, or raise the water table, or measure the drainage water volume when the water table is lowered. Most percolation lysimeters use manual vacuum pumping or volumetric collection to measure drainage. Some weighing lysimeters that have greater belowground access have used tipping-bucket devices or tanks suspended from load cells from the lysimeter itself to measure drainage rates or volumes.

The main difference between percolation and weighing lysimeters is that weighing lysimeters measure the mass of the lysimeter (in megagrams or kilograms), and this mass change is then converted, taking into account the lysimeter area (meters squared) and the density of water (typically assumed to be $1\,\mathrm{Mg\,m^{-3}}$; note: $1\,\mathrm{kg\,m^{-2}} = 1\,\mathrm{mm}$ assuming a water density of $1\,\mathrm{Mg\,m^{-3}}$), to determine θ_{i}. Assuming no drainage, precipitation, or irrigation during the measurement time period ($\mathrm{ET}_{\mathrm{i}} = \theta_{\mathrm{i}} - \theta_{\mathrm{i-1}}$), the lysimeter mass measurement precision is a direct ratio of the lysimeter area (i.e., $1\,\mathrm{kg\,m^{2}}$ per millimeter), while the soil volumetric measurement precision is much greater (i.e., $0.001\,\mathrm{m^{3}\,m^{-3}}$ per millimeter). Therefore extremely high soil-water measurement precision is required to determine ET accurately for time periods of a single to a few days. Weighing lysimeters with proper design and electronic measurements can achieve equivalent accuracies for a day without counterbalancing. With counterbalancing of the soil 'dead' mass, excellent ET measurement precision has been obtained for time periods from 15 min to 1 h with typical weighing lysimeters. Weighing lysimeter mass measurement precisions have been noted to decline as wind speeds exceed approximately $3\,\mathrm{m\,s^{-1}}$ due to wind load imposed on the lysimeter. Under very windy conditions, longer mass integration intervals may be necessary or numerical smoothing can be used.

Lysimeter Design Considerations

Experimental purpose is the principal consideration in lysimeter design. If the lysimeter facility is intended to represent natural outdoor environmental conditions, then siting becomes a most critical consideration. The site should represent the soil and plant conditions desired. Surrounding conditions, for perhaps up to a kilometer or more, can certainly impact the use of water by a crop, especially for taller crops. It is important to be aware that in many cases these surrounding areas may not be directly controlled by the investigators using the lysimeters, and features and land use of surrounding land areas can change rapidly, compromising the lysimeter facility. Siting needs to include considerations for adequate surrounding fetch (similar cropped distance usually expressed in the direction of the predominant wind direction), utility access (telephone, electrical power, and sometimes water), and foot-traffic access for maintenance. Fetch needs vary with the research purpose and site conditions, but easily extend from 100 to 200 m. Utility access needs vary if alternate current (AC) power is required or if a telephone modem is used for data transfer. Most current lysimeter data loggers are battery-powered and some use an AC power backup for the batteries. Although analog telephone lines and modems are commonly used for data transfer, cellular telephones are becoming more common, as well as other wireless network technology. All lysimeters require repeated, periodic site visits for maintenance and measurements. Care must be taken to avoid crop damage and 'path' creation that might affect the lysimeter microclimate.

Lysimeter Shape and Area

Percolation and weighing lysimeter shapes are typically circular, square, or rectangular. Cylindrical tanks are inherently stronger per unit wall thickness than square or rectangular tanks, owing to their greater hoop strength. However, circular lysimeters can cause inaccuracies, depending on the vegetation type, especially for row crops with wide row spacing.

Figure 2 shows several typical lysimeter tank surface configurations. An important consideration is the lysimeter surface area. Obviously, larger sizes are desired, but they are constrained by capital resources as well as physical ones. Lysimeter areas in use vary considerably from less than 1 to 30 m^2 or greater. Two principles guide the selection of the appropriate area: (1) soil pedological considerations (size needed to duplicate saturated hydraulic conductivity of natural soil; i.e., macropores and soil cracks); and (2) the unit area of vegetation that might be considered as a minimum representative sample. The soil-pedology factor is particularly important for more structured, heavy vertisols and mollisols that have a high level of shrink–swell and structural cracking. The vegetation-area factor is particularly important in terms of matching the row spacing of crops or the size of the area needed to capture the natural variation in native species. This area is linked to the balance of the evaporative processes for soil water (E) and plant transpiration (T). For closely seeded crops (alfalfa, small grains, grasses, etc.) or selected forestry or native brush species, circular shapes are adequate; but for agronomic crops or horticultural crops (trees, vines, or vegetables) grown in rows, a square or rectangular shape is desirable. Figure 2 shows the 'equivalent' surface area of small, circular lysimeters, which introduces an uncertainty in knowing exactly the 'effective area' of the lysimeter. The square area of the row or multiple rows through the lysimeter (see the dashed lines in Figure 2) represents the sunlight-absorption area of the crop. The E and T components of a circular shape are constrained by the lysimeter surface area. For larger circular lysimeters with many rows, the lysimeter diameter can depart from an even multiple of the row spacing without a large bias on the balance of E and T and not affect the measured ET accuracy. However, for smaller circular lysimeters with a single row or even two rows, it is difficult to avoid biasing the ET accuracy by unbalancing the E and T components of crop water use if the lysimeter diameter is not larger than the row spacing (i.e., for a single row, row spacing is less than lysimeter diameter, which is less than 1.8 × row spacing; for a single row, this is $D_l = \pi^{-1/2} W$, where W is row spacing (in meters) and D_l is lysimeter diameter (in meters)).

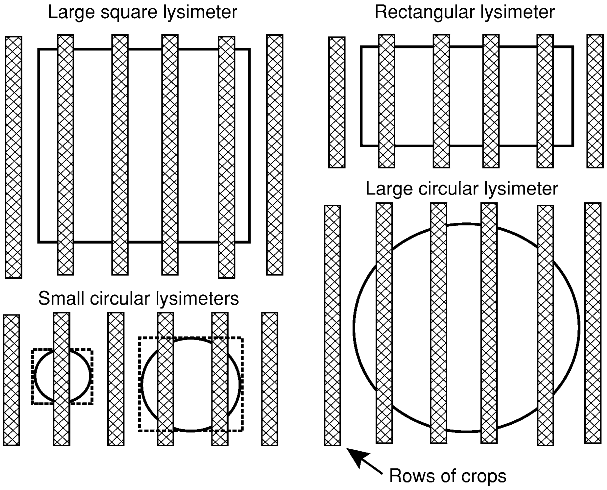


Figure 2 Lysimeter shapes for large and small lysimeters. The dashed lines represent the 'equivalent' area of sunlight interception for the smaller circular lysimeters.

An often unreported factor for interpreting lysimeter water balances has been the lysimeter effective area. For percolation lysimeters and some weighing lysimeters, the area considerations discussed above are important. Most percolation lysimeters have a single wall (Figure 1), so the effective area is assumed to match the inside tank area (i.e., $\pi\ (D_l^2/4)$ for a circular lysimeter and L_l^2 or $L_l \times W_l$ for a square or rectangular lysimeter tank, where L_l and W_l are the lysimeter length and width, respectively, for a square and rectangular design). For a weighing lysimeter, which has two walls (the inner tank and outer tank walls) and an air gap (see Figure 1), the effective area for evaporation is computed for the 'midgap' dimensions. This midgap and wall area might represent approximately 5–10%, or even greater in some cases, of the lysimeter tank inner area. So the measured change in mass is adjusted downward for this area over which the outside and inside vegetation are assumed to be equal. Some hydrologic weighing lysimeters have used precisely beveled upper gap covers to capture the precipitation on the lysimeter correctly.

Lysimeter Depth

The depth of a lysimeter is an important design consideration; greater depths increase construction and installation problems. The basic consideration is to duplicate adequately the field rooting depth for the species being studied. Additionally, lysimeter depth has been shown to affect the lysimeter water-holding capacity and the soil-water flux through the profile. If the lysimeter depth is less than 1.5–2 m, vacuum drainage is required to mimic natural soil-water pressure potentials. If a water table is used to match field conditions and/or to simulate high water tables, care must be exercised in selecting the desired lysimeter depth. Few lysimeter designs have used depths greater than 2.5–2.7 m; but several newer lysimeters (one in China) are 5 m in depth, while some in Arizona in the western USA are 4 m in depth. For weighing lysimeters, greater depth (volume per unit area) necessitates counterbalancing a large fraction of the soil and tank mass to achieve high precision in the ET measurements.

Soil Profile Characteristics

The lysimeter soil profile should match or duplicate the site characteristics so that vegetative growth and water use will be representative of the site and experimental purposes. Both soil monoliths and repacked (reconstructed) soil profiles are used. Under 'well-irrigated' cropping situations, a repacked soil lysimeter column is often adequate. Exact physical, chemical, and/or vegetation characteristics of a soil can only be preserved in many cases by using a soil monolith. Naturally occurring soil spatial heterogeneity (both vertically and horizontally) can affect soil hydraulics, chemical composition, and thermal processes. In many lysimeters, repacked soil profiles have been successfully used for coarsetextured soils and finetextured soils that do not have significant profile stratification. Larger soil monoliths are more difficult to obtain and often require heavy industrial-type equipment (cranes, backhoes, coring equipment, etc.), but the time required to obtain a representative-sized monolith (approximately 5–15 m^3) must not exceed the time or labor needed to repack the same soil volume. Also, an intact soil monolith is ready for research applications almost instantly following collection, while repacked soil columns might take several years to duplicate the natural soil state for research with restricted irrigation (i.e., rain-fed or dryland applications). Figure 3 shows a lysimeter being repacked.

Figure 3 Technician installing soil-water sensors in a repacked lysimeter near Five Points, California. The lysimeter is a 2-m square (4 m^2). (Photo courtesy of USDA–ARS, Water Management Research Laboratory, Parlier, California, USA.)

Soil monolith collection Large soil monoliths are usually collected by coring, encasing, or compositing several smaller monoliths together. In the coring method, an open-ended tank (usually steel) is pushed into the soil using static weights or pulled into the soil using an anchor secured to the earth and various hydraulic or manual jacking methods. The tank's leading edge may be beveled to facilitate soil shearing while reducing internal compression on the monolith. In most cases, coring requires outside excavation as the monolith tank is lowered to reduce outside soil friction and to reduce the shearing forces. Soil monolith tanks that are square or rectangular are supported on the outside of the tank walls, while similar tanks that are repacked can be structurally supported by the floor and internal gussets as well as horizontal wall supports. Circular monolith tanks require simpler hoop-bracing. Figure 4 shows a square soil monolith prior to being pressed into the soil using four anchors to earth piers using a manually operated, hydraulic jacking assembly (shown on top of the lysimeter tank). This hydraulic lowering process allows precise control of installation speed in contrast to adding static 'dead' weights. The hydraulic installation

Figure 4 A monolith soil tank (1.5 m square and 2.3 m in depth) ready for insertion at Bushland, Texas. The hydraulic rams are located on the top corners and offset to permit a backhoe to excavate along the outside of the tank as it is pressed into the soil against the anchors secured into concrete piers. (Photo courtesy of USDA–ARS, Conservation and Production Research Laboratory, Bushland, Texas.)

procedure using rams or jacks avoids any erratic, jerky loading such as those that can occur when static weights are added.

Soil monoliths have been obtained in structured soils by carefully trimming a soil pedestal and gradually pressing a container tank over the soil pedestal while trimming the leading edge or using expandable foam (care has to be used in material selection as some materials react with or leach chemicals into the soil) to seal the soil pedestal against the tank walls. In encasing, a trimmed soil pedestal is carefully fitted with sections that can later be sealed to form the tank. In some encasings, pylon-like sheets are driven into the soil and welded or sealed together after excavating around the soil monolith.

In both coring and encasing methods, the soil monolith is undercut and lifted before a drainage system and tank bottom is installed. Undercutting methods are varied, too, including forcing shear plates under the monolith, boring holes and using 75- to 100-mm pipes, and driving steel rods or structural steel (angle iron, channel iron-type cross sections) to support the monolith during lifting. The boring method works best with more structured subsoils, while plates may be required for coarse sand or gravel subsoils. Monoliths may be desirable in certain situations to preserve natural soil-profile layering that cannot be reconstructed; but they are primarily soil samples with unique heterogeneities inherent in the spatial soil variability at the site. If soil monoliths have to be transported to the lysimeter site, care must be used to minimize the vibrations that inevitably occur during transport which can physically damage the monolith.

Lysimeter tanks The lysimeter tank, whether for a repacked column or for a monolith, should include measures to reduce water flow down the inside walls and even perhaps a division in the tank floor to separate the wall drainage from the inner core drainage. In repacked lysimeter tanks, the inside wall flow can be buffered by the inside horizontal bracing. In monolith tanks, an antiseep collar may be required at a short distance (approximately 200–300 mm) below the soil surface to reduce direct wall flow, particularly for a soil that may shrink away from the inner tank walls. Larger lysimeters require an adequately designed floor-bracing system to avoid tank-floor warping and for scale supports for weighing lysimeters.

Lysimeter drainage systems Lysimeter drainage system designs are quite varied, from those used to maintain static or falling water tables, to those with gravity drainage or manually pumped vacuum drainage, to those with automated vacuum systems. Gravity drainage can be measured volumetrically or by rate (tipping-bucket types). Manual vacuum drainage is often measured volumetrically while pumping. Automated vacuum systems have been used to maintain a constant vacuum on drainage lines, with drain outflows into tanks attached to a weighing lysimeter or suspended with load cells. Drain materials vary widely also. Gravity drains are made from simple, perforated polyvinyl chloride (PVC) pipes or perforated polyethylene (PE) pipes. Vacuum drains are made from ceramic cups, porous blocks, or sintered stainless tubes; and it is important that the 'bubbling pressure' of the soil in contact with the drain media is not exceeded. Some lysimeter drain systems have used outside soil matric potential sensors to match the drainage vacuum and even to add water to the lower soil profile if upward fluxes are detected. The drainage lines are usually embedded in fine sand or diatomaceous earth to apply the vacuum to the soil column uniformly.

Weighing-Lysimeter Scale Systems

Lysimeter scale systems are highly diverse and often uniquely conceived by the lysimeter design team. Scale systems include commercial lever load-cell types, flexure load cells, hydraulic floating-cell types, and hydraulic load-cell types. Mechanical lever-type scales permit a large part of the soil 'dead' (dry) mass and the tank mass to be tared statically to permit a small- to medium-sized electronic load cell to be used to measure the 'live' mass changes (water). Some mechanical scales have movable tare weights to adjust the offset counterbalance mass to maintain a small electronic measurement range. The greatest advantages in weighing lysimetry are probably widespread availability and low-cost, highly precise analog-to-digital data loggers now commonly used to measure load cells. Typical ET mass accuracy can easily surpass 0.05- to 0.02-mm ranges. Modern data loggers also permit intensive data collection to monitor load-cell drift and other possible electronic problems.

Floating lysimeters were popular before the more expensive electronic–mechanical scales became popular with the improved data-recording precision. Floating lysimeters were designed with lightweight ballast tanks beneath the lysimeter tank, so highly accurate recordings (usually manual liquid depths that were later measured electronically) could be made of the lysimeter displacement to track the lysimeter mass changes. Often the more sensitive of these floating lysimeters were even more accurate than the mechanical scale types, with accuracies as small as 0.025 mm.

Various hydraulic weighing lysimeters have been built with a diverse range of hydraulic reading mechanisms. Some have used simple, homemade hydraulic load cells (often referred to as 'pillows') to measure the lysimeter mass change as the load (pressure) on the hydraulic load cells changes. Hydraulic weighing lysimeters are subject to a greater uncertainty owing to the density changes of the fluid (usually water or a glycol–water solution) and/or a Hg manometer due to temperature changes or flexure changes in the 'pillow' materials. These hydraulic weighing lysimeters achieved accuracies approaching 0.05–0.1 mm.

Lysimeters have been used with only load cells or load-beam scales without counterbalances. Placement of the load cells or scale platform under the lysimeter tank (at approximately 1.5- to 2.5-m depths) greatly reduces temperature influences on the electronic load cells. With highly precise data loggers and larger area-to-mass lysimeters, accuracies within the range 0.1–0.05 mm have been reported. These accuracies are acceptable for good short-term ET research, while greatly reducing the lysimeter scale complexity and construction costs.

Measuring ET or Latent Heat Fluxes with Weighing Lysimeters

The ET rate or mass ET accumulation can be measured easily with precise lysimetry provided no water is added or drained from the lysimeter during the measurement period. If water additions or drainage occur, the water balance can be calculated, but the additions or drainage must be computed separately to determine ET (see eqn [1]). This is a common problem with lysimetry in hydrology. Seldom will either the 'gross' precipitation or irrigation volumes be known in accurate comparison with the measured net soil mass water balance determined via lysimetry. During periods without water additions or drainage, the ET rate is computed as:

$$\mathrm{ET_i} = \{A_\mathrm{l}[(M_\mathrm{i} - M_\mathrm{i-1})/A_\mathrm{l}]/A_\mathrm{f}\}/T_\mathrm{i} \qquad [2]$$

where: ET is the evapotranspiration rate, in millimeters per hour (kilograms per square meter per hour), for time interval *i*; M is lysimeter mass, in kilograms; A_l is the lysimeter inner tank surface area, in square meters; A_f is the lysimeter foliage area (mid wall–air gap area), in square meters; and T is the time in period *i* in hours. The ratio A_f to A_l^{-1} is the correction factor for the lysimeter effective area. This correction factor assumes the outside and inside vegetation foliage overlap evenly on all of the sides or edges. If there is no overlap, as might occur with a shorter grass, then the correction factor becomes 1.0. Various techniques have been used to obtain reliable short-term ET rates. The lysimeter depletion over a short interval (approximately 30 min to 1 h) might be plotted and a linear line fit to the data so the derivative (ET rate) can be computed for that period. Alternatively a 'numerical' difference can be computed from eqn [2]. If the M terms are 'noisy' (common with high winds), a numerical smoothing technique can be used (equal or weighted averaging running averages).

The latent heat flux is λET, where λ is the latent heat of vaporization in megajoules per kilogram. When the time interval is in seconds, then λET will be in watts per meter squared. Figure 5 shows the ET rate of irrigated alfalfa on 1998 June 13–14 (day of year (DOY) 164 and 165) at Bushland, Texas, measured with two precision weighing lysimeters. On DOY 164, advection took the form of exceptionally strong winds and a large vapor pressure deficit (VPD), while the second day had calmer winds and a lower VPD, with almost identical net radiation. The increased wind speeds on the 2nd day during the 18.00–19.00 time period was noticeable together with the greater postmidnight, predawn evaporation on that day; while winds and VPD were still greater than those typically experienced in that time period. On DOY 164 peak λET exceeded 1000 W m^{-2} (greater than peak incoming solar irradiance), but, because the crop had a fetch of almost 300 m (NE lysimeter; 100 m fetch for the SE lysimeter) and was 'well watered,' it could maintain the extremely large transpiration fluxes with no discernible 'stomatal control' above typical evaporative conditions (i.e., the next day).

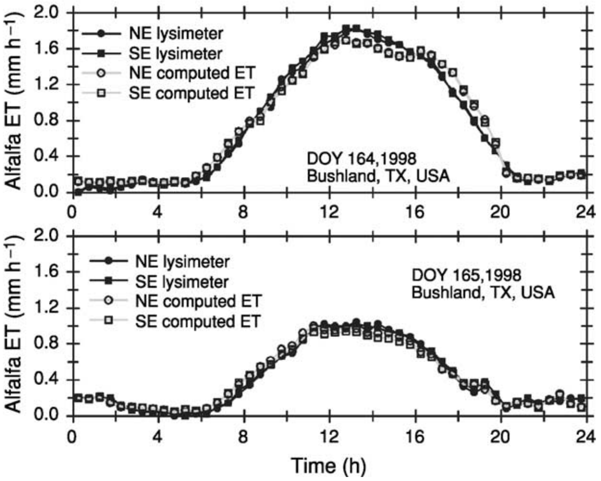


Figure 5 Example alfalfa evapotranspiration (ET) measured with two precision lysimeters at Bushland, Texas, under advective conditions (DOY 164) and on a more typical day (DOY 165) in 1998 contrasted with reference alfalfa ET computed with a Penman–Monteith equation and standardized parameters based on the ASCE/EWRI (American Society of Civil Engineers/Environmental Water Resources Institute) procedures.

Summary

Lysimeters are hydrologic instruments to measure the percolating water quality and/or to measure the evaporation of water from vegetated surfaces. They can be percolation or weighing types. The water balance is measured in percolation types by soil water balance while the mass change in the weighing types measures the water use (or gain from irrigation or precipitation). The lysimeter design must carefully consider its purpose. Then its size, depth, type, and drainage are key considerations. With well-designed lysimeters that are carefully maintained, high-quality data can be obtained on vegetation water use and percolation water quality.

See also: **Drainage, Surface and Subsurface**; **Evaporation of Water from Bare Soil**; **Evapotranspiration**; **Leaching Processes**; **Macropores and Macropore Flow, Kinematic Wave Approach**; **Neutron Scattering**; **Salination Processes**; **Salinity:** Management; **Soil–Plant–Atmosphere Continuum**; **Solute Transport**; **Tensiometry**; **Time-Domain Reflectometry**; **Vadose Zone:** Hydrologic Processes; **Water Content and Potential, Measurement**; **Water Cycle**; **Water-Use Efficiency**

Further Reading

Aboukhaled A, Alfaro A, and Smith M (1982) *Lysimeters*. FAO Irrigation and Drain Paper No. 39. Rome, Italy: Food and Agricultural Organization of the United Nations.

Allen RG and Fisher DK (1990) Low-cost electronic weighing lysimeters. *Transactions of the American Society of Agricultural Engineers* 33: 1823–1833.

Allen RG, Howell TA, Pruitt WO, Walter IA, and Jensen ME (eds) (1991) *Lysimeters for Evapotranspiration and Environmental Measurements*. Proceedings of the International Symposium on Lysimetry. New York: American Society of Civil Engineers.

Bergström L (1990) Use of lysimeters to estimate leaching of pesticides in agricultural soils. *Environmental Pollution* 67: 325–347.

Brutsaert W (1982) *Evaporation into the Atmosphere*. Dordrecht, The Netherlands: D. Reidel.

Duke HR and Haise HR (1973) Vacuum extractors to access deep percolation loses and chemical constituents of soil water. *Soil Science Society of America Proceedings* 37: 963–964.

Harrold LL and Dreibelbis FR (1951) *Agricultural Hydrology as Evaluated by Monolith Lysimeters*. Technical Bulletin No. 1050. Washington, DC: US Department of Agriculture.

Hatfield JL (1990) Methods of estimating evapotranspiration. In: Stewart BA and Nielsen DR (eds) *Irrigation of Agricultural Crops*, pp. 435–474. Monograph No. 31. Madison, WI: American Society of Agronomy.

Howell TA, Schneider AD, and Jensen ME (1991) History of lysimeter design and use for evapotranspiration measurements. In: Allen RG, Howell TA, Pruitt WO, Walter IA, and Jensen ME (eds) *Lysimeters for Evapotranspiration and Environmental Measurements*, pp. 1–9. Proceedings of the International Symposium on Lysimetry. New York: American Society of Civil Engineers.

Kirkham RR, Gee GW, and Jones TL (1984) Weighing lysimeters for long-term water balance investigations at remote sites. *Soil Science Society of America Journal* 48: 1203–1205.

Marek TH, Schneider AD, Howell TA, and Ebeling LL (1988) Design and construction of large weighing lysimeters. *Transactions of the American Society of Agricultural Engineers* 31: 477–484.

Pruitt WO and Angus DE (1960) Large weighing lysimeter for measuring evapotranspiration. *Transactions of the American Society of Agricultural Engineers* 3: 13–15, 18.

Ritchie JT and Burnett E (1968) A precision weighing lysimeter for row crop water use studies. *Agronomy Journal* 60: 545–549.

Schneider AD and Howell TA (1991) Large, monolithic, weighing lysimeters. In: Allen RG, Howell TA, Pruitt WO, Walter IA, and Jensen ME (eds) *Lysimeters for Evapotranspiration and Environmental Measurements*, pp. 37–45. Proceedings of the International Symposium on Lysimetry. New York: American Society of Civil Engineers.

Schneider AD, Howell TA, and Steiner JL (1993) An evapotranspiration research facility using monolithic lysimeters from three soils. *Applied Engineering in Agriculture* 9: 227–232.

Schneider AD, Steiner JL, and Howell TA (1994) Pulldown forces for collecting large soil monoliths. *Transactions* *of the American Society of Agricultural Engineers* 37: 831–835.

Schneider AD, Howell TA, Moustafa ATA, Evett SR, and Abou-Zeid WS (1998) A simplified weighing lysimeter for monolithic or reconstructed soils. *Applied Engineering in Agriculture* 14: 267–273.

Soileau JM and Hauck RD (1987) A historical review of US lysimetry with emphasis on fertilizer percolation losses. In: *Proceedings of the International Conference on Infiltration Development and Application.* St. Joseph, MI: American Society of Agricultural Engineers.

Tanner CB (1967) Measurement of evapotranspiration. In: Hagan HM, Haise HR, and Edminister TW (eds) *Irrigation of Irrigated Lands*, pp. 534–574. Monograph No. 11. Madison, WI: American Society of Agronomy.

Yang J, Li B, and Liu S (2000) A large weighing lysimeter for evapotranspiration and soil-water-groundwater exchange studies. *Hydrologic Processes* 14: 1887–1897.

Young MH, Wierenga PJ, and Mancino CF (1996) Large weighing lysimeters for water use and deep percolation studies. *Soil Science* 161: 491–501.

M

MACRONUTRIENTS

C W Wood, J F Adams, and B H Wood,
Auburn University, AL, USA

Introduction

Macronutrients are defined as plant nutrients that typically attain a concentration $>500\,\mathrm{mg\,kg^{-1}}$ in mature plants. They include nitrogen (N), phosphorus (P), potassium (K), calcium (Ca), magnesium (Mg), and sulfur (S). Proper management of these nutrients is essential to ensure that farmers get maximum economic yield while protecting environmental quality. Sensible management of N and P by producers is especially important for environmental protection, because these elements ordinarily limit plant growth in natural terrestrial and aquatic ecosystems. Eutrophication (enrichment) of natural ecosystems with N and P can result in a myriad of undesirable environmental consequences, such as loss of species diversity, fish kills, increased cost of water treatment, impaired navigational and recreational use of waterways, noxious odors, and increased populations of insect pests.

Conservation management of macronutrients implies that these elements will be handled in a manner promoting incorporation into harvestable plant biomass and retardation of movement away from the site of application. To accomplish this goal, management decisions should be based on their cycling patterns, which vary among macronutrients. Two macronutrients (N and S) have gaseous pathways in their cycles that complicate management, while the others have simpler cycles with no atmospheric pathway.

Nitrogen

Nitrogen is the most limiting nutrient in biomass production systems, due to its large requirement by plants and the finite ability of soils to supply available N. Plants take up N as nitrate (NO_3), ammonium (NH_4), or urea ($CO(NH_2)_2$). In moist, warm, well-aerated soils the NO_3 form is dominant. Once inside plants NO_3 is reduced to NH_4 using energy provided by photosynthesis. Plants normally contain between 10 and $50\,\mathrm{g\,N\,kg^{-1}}$ on a dry basis.

Conservation of N in biomass production systems involves preventing losses of N that enters soils through various N-cycle pathways. These pathways include: (1) addition by precipitation; (2) fixation by lightning; (3) biological N fixation, both symbiotic and asymbiotic; (4) addition of organic fertilizers and crop residues; and (5) addition of industrially fixed N as fertilizers. Nitrogen added to soil systems in precipitation is generally in the form of NO_3 and NH_4, and usually amounts to $5\text{–}10\,\mathrm{kg\,N\,ha^{-1}\,year^{-1}}$, with greatest amounts being deposited in regions where acid rain is prevalent. Fixation by lightning occurs when it oxidizes atmospheric dinitrogen (N_2) to $NO + NO_2$, which in turn is converted to NO_3, but this only amounts to 10–20% of the N deposited via precipitation. Amounts of N added in precipitation are not enough to support agricultural biomass production, but may be enough to sustain mature natural ecosystems.

N Fixation

Atmospheric N_2 can be fixed asymbiotically or symbiotically by specialized organisms into organic forms that can subsequently be mineralized into plant available forms. Asymbiotic fixation is performed by free-living organisms such as blue-green algae that fix up to $25\,\mathrm{kg\,N\,ha^{-1}\,year^{-1}}$, and aerobic (e.g., *Azobacter* spp.) and anaerobic (e.g., *Clostridium* spp.) bacteria that fix up to $0.5\,\mathrm{kg\,N\,ha^{-1}\,year^{-1}}$. These organisms do not require a specific host, unlike those which bring about symbiotic N fixation. The latter include rhizobia bacteria spp. in legumes and actinomycete spp. in some trees, that fix atmospheric N_2 in symbiotic association with the higher plants. These microorganisms reside in nodules on plant roots and derive their energy from carbohydrates produced photosynthetically by their higher plant symbiont. Symbiotic fixation's contribution of N to terrestrial biomass production systems can be substantial. For example soybean, cowpea, clover, lupin and alfalfa have been shown to fix up to 75, 85, 130, 160, and

$360\,\mathrm{kg\,N\,ha^{-1}\,year^{-1}}$, respectively. The alder, a small tree, has been shown to fix up to $170\,\mathrm{kg\,N\,ha^{-1}\,year^{-1}}$. Symbiotic N fixation has been recognized as being second in importance among biological processes only to photosynthesis in regard to sustaining life on earth.

Fertilizers and Manures

Atmospheric N_2 is fixed industrially as NH_3 via the Haber-Bosch process, which requires high temperature (1200°C) and pressure (350 atm). Subsequently the NH_3 gas can be converted to various liquid and solid NO_3, NH_4, or $CO(NH_2)_2$ fertilizers.

The major types of organic fertilizers include barnyard manures, liquid manures, processed organic materials, and crop residues returned directly to the soil. Barnyard manures are relatively dry, bulky, solid materials that derive their nutrient value from animal excreta. Liquid manures consist of animal excreta that have been deposited on solid or slatted floors, without bedding material, and then washed into lagoons or storage tanks. Processed organic materials are produced on- or off-farm and include composts, sewage sludge, food-processing wastes, forestry by-products, industrial wastes, and municipal refuse. Crop residues include plant parts (straw, stover, roots, etc.) that remain on the land after crop harvest and green manure or cover crops grown for incorporation into the soil.

Organic fertilizers vary widely in N concentration depending on their origin, and consequently, some materials have greater N fertilizer value than others. Of the barnyard manures, those generated by birds tend to have the greatest N concentration owing to their relatively low water content. Due to low dry matter concentrations, liquid manures typically have low N concentrations. Sewage sludges tend to have relatively high N concentrations. N concentration generally declines when organic wastes are composted, but co-composted dead poultry and chicken manure may have a greater N concentration than barnyard manures. Crop residues (e.g., maize and cotton stover, and wheat and rice straw) usually have low N concentrations in comparison to leguminous green manures (e.g., clover and alfalfa) which, due to their typically greater N concentration, can provide substantial amounts of residual N for succeeding crops.

Typically, for net release of plant available N to occur in the soil, i.e., net mineralization, organic fertilizers should have a carbon (C):N ratio <20. Application of organic materials with C:N >30 will result in immobilization of N by soil microbes. Typically, for immediate release of plant available N, organic fertilizers should have a N concentration $>17\,\mathrm{g\,kg^{-1}}$ on a dry-weight basis. The end product of the mineralization of organic forms of N is NH_4, which can be further transformed to NO_3 via the process of nitrification.

Nitrogen-cycle pathways that lead to soil system losses include crop removal, leaching, runoff, NH_3 volatilization, biological denitrification, chemical denitrification, and urea hydrolysis. From a biomass production standpoint, crop removal is the desired outcome, while the others result in economic loss to producers and pollution of the environment.

Leaching results in movement of NO_3 below the rooting zone. Nitrate that escapes the rooting zone can eventually enter groundwater and present a health hazard to humans and livestock. Management to prevent nitrate escape into groundwater includes use of appropriate N application rates; losses of NO_3 via leaching will be minimized if N is not added beyond that needed for maximum yield. Synchronization of N application with crop needs will also decrease risk of NO_3 leaching; plants need little N at planting when the potential for leaching loss can be high. Split or multiple applications in irrigation water allow plants to be fertilized when they need N most, lowering potential for leaching losses. Slow-release N fertilizers may also reduce NO_3 leaching losses, but are expensive, and generally limited to use on high-value crops such as vegetables and turf. Lastly, winter cover crops can help reduce NO_3 leaching by capturing NO_3 that otherwise would be lost below the rooting zone.

Losses of N in runoff are a function of rainfall amount and intensity, infiltration of water into soil, and vegetative cover. Management to prevent losses of N in runoff include maintaining vigorous ground covers on production fields when possible, and maintaining healthy filter strips and riparian areas on borders of production fields year-round. Any management that results in greater water infiltration into the soil will decrease runoff N losses, but may increase losses via leaching and/or denitrification.

Under alkaline conditions the NH_4 ion is unstable and subject to chemical decomposition and subsequent volatilization: $NH_4{:}NH_{3(solution)} + H$; $NH_{3(solution)}{:}NH_{3(gas)}$. Volatilization losses from surface-applied N can be substantial ($>50\%$ of N applied). Ammonia volatilization represents an economic loss to farmers, and atmospheric NH_3 can promote negative environmental impacts such as respiratory problems, acid rain production, soil acidification and ecosystem eutrophication following deposition, and toxicity to plant tissue. Many factors promote NH_3 volatilization as they increase: surface area of the soil solution–atmosphere interface, velocity of air moving across the soil surface, temperature, partial pressure of NH_3 in soil solution, and pH.

Ammonia is also volatilized when surface-applied urea, $CO(NH_2)_2$, comes in contact with the ubiquitous soil enzyme urease. Reduction of NH_3 volatilization is most easily accomplished via techniques such as tillage, injection (liquid manures), irrigation, or timing of application with natural rainfall, which move NH_3/NH_4 to depths where it can be adsorbed by soil colloids.

Biological denitrification is the facultative microbial reduction of NO_3 and nitrite (NO_2) with liberation of N_2 and nitrous oxide (N_2O). Common, widespread species of typically aerobic bacteria use NO_3 and NO_2 as an oxygen source in the absence of O_2. The potential for denitrification losses is high given the appropriate requisites including: anaerobiosis, carbohydrates for energy, $pH > 5.5$, NO_3/NO_2 supply, and warm temperatures. Besides being an economic loss to farmers, release of N_2O into the atmosphere via biological denitrification serves to exacerbate the 'greenhouse effect' by decreasing escape of infrared energy into space. Moreover, N_2O has been implicated in photo-oxidation of stratospheric ozone (O_3), which increases penetration of destructive ultraviolet rays to the earth's surface. Managing for inhibition of biological denitrification involves any processes that: (1) slow nitrification, such as chemical nitrification inhibitors; (2) enhance NO_3 removal such as plant uptake; and (3) promote aeration, such as drainage and evapotranspiration.

Denitrification can also occur chemically under aerobic conditions when NH_4 and NH_4-forming fertilizers are nitrified. This result can be realized through different chemical pathways, but they all involve destruction of the nitrification intermediate NO_2. Thus, from a management standpoint, chemical denitrification is not an important process since NO_2 is ephemeral in soil systems.

Phosphorus

Phosphorus is one of the key essential elements in modern agriculture, and it has many important functions in plants, the premier one being energy storage and transfer. The agronomic literature is replete with examples of grain, fiber and forage yield increases owing to proper maintenance of P fertility in arable soils. Although salutary effects of P in agricultural production are evident, P can also function as a pollutant if it moves from the site of its intended use. The main concern is P transport in surface or subsurface flow from soils to streams, rivers, lakes, and eventually the oceans. Phosphorus transported from agricultural soils can promote eutrophication, and algal growth in surface waters is usually limited by P instead of N availability.

Origins of P in soils include residual soil minerals, or inputs of P from commercial fertilizers and organic fertilizers/wastes. The most common mineral forms of P in soils are: (1) apatite (calcium phosphate), which is found in unweathered and moderately weathered soils; and, (2) iron and aluminum phosphates, which are found in highly weathered soils. The raw material for commercial fertilizer P is apatite mined from various deposits around the world and treated with acid to increase its solubility. Organic P fertilizers/wastes derive from a variety of sources including animal manures, crop residues, and municipal and industrial wastes.

Once a P atom enters the soil environment it is subject to several soil processes that control its fate. In soil solution, P is present as either a monovalent (H_2PO_4) or divalent (HPO_4) anion, the form depending on soil pH. Phosphorus enters the soil solution via either: (1) dissolution of primary minerals; (2) dissolution of secondary minerals; (3) desorption of P from clay and mineral surfaces; and (4), biological conversion of P in organic materials to inorganic forms (mineralization). In most soils, soil solution P ranges between <0.01 and $1\,mg\,l^{-1}$, and a value of $0.2\,mg\,P\,l^{-1}$ is commonly accepted as the solution P concentration desired to support agronomic crops.

Dissolution and erosion are the processes by which P is moved into runoff from agricultural fields. Dissolution of P from a thin zone of surface soil and vegetative material yields dissolved P that is immediately available for uptake by aquatic biota. Erosion via surface runoff transports particulate P in the form of soil and vegetative matter; only that portion of particulate P in equilibrium with dissolved P is available for aquatic biota. Thus, bio-available P includes dissolved P and a portion of particulate P. Once bio-available P moves from agricultural fields into receiving waters it can contribute to eutrophication.

Although generally considered a less important mechanism than surface runoff, P leaching followed by shallow lateral subsurface flow can contribute dissolved P to surface waters under high water table conditions. This mechanism becomes more important in soils with large accumulations of P that saturate surface soil sorption capacity, leading to downward movement of P.

Escape of soil P to surface waters is affected by a host of soil management practices (e.g., tillage method, cover crops, riparian zones, terracing, etc.). However, the soil P concentration that correlates with P bioavailability is the greatest determinant of the balance between adequate soil P fertility and off-site P escape. In that regard, soil testing is currently the best management tool available to ensure that crops are provided with adequate soil P. Moreover, soil testing is likely the best management tool available to ensure that soils do not become overloaded with P, which increases the likelihood of their contribution to pollution of downstream water.

Potassium

Potassium is the seventh most common element in the earth's crust, accounting for 2.6% of its mass. It is present in igneous, sedimentary, and metamorphic rocks, and deposits that are mined for fertilizer are evaporites, largely KCl, derived from drying of enclosed bodies of seawater. The quantity of K in soils is dependent on parent material and degree of weathering. Minimally weathered soils in arid regions may contain up to 2.5% K, while weathered soils in the tropics may contain little or no K.

Potassium is taken up by plants in amounts greater than those of N, and depending on species, plant tissue concentrations can range as high as 5%. It does not combine with C in plants to form organic compounds, but exists in a mobile ionic form (K^+). Potassium has many functions in plants, but owing to its high tissue concentration and ionic form, it is not surprising that its primary one is osmoregulation.

Potassium has no atmospheric pathway, and exists in the soil as primary minerals (feldspars and micas), secondary minerals (illite and vermiculite), exchangeable K, and soil solution K. Of the primary minerals, feldspars are more resistant to weathering than micas (biotite and muscovite), and exist mostly in the sand and silt size fractions. Weathered micas exist in the clay size fraction, and eventually release K from interlayer spaces as they hydrate to form secondary minerals. In a reverse process, K can be trapped when the interlayer spaces of secondary minerals collapse upon dehydration, reforming mica. When K^+ leaves the interlayer space of a secondary mineral, it will likely be adsorbed on a cation exchange site of a soil colloid. This exchangeable K maintains an equilibrium with the K^+ in soil solution, together constituting 'available K.'

Loss pathways for K include leaching and erosion/runoff. Leaching of K is increased under conditions that include a high degree of soil colloid saturation with K, sandy subsoils, poor vegetative cover and shallow rooting, and low soil-pH. Soil acidity impacts K leaching through the complementary ion principle: there will be less leaching of K if Ca^{++} rather than hydrogen (H^+) or aluminum (Al^{3+}) is the dominant complementary ion on soil-colloid exchange sites. Thus, management to reduce K leaching from soils includes following soil test recommendations to avoid saturation of soil colloids with K, maintaining vigorous vegetative cover, and liming to maintain soil pH according to soil test recommendations. In regard to K loss via erosion/runoff, management previously described to prevent surface soil losses of P will also prevent K loss.

Calcium

Calcium concentration in the earth's crust is approximately 3.6%, while soils contain only 1.37%. The primary source of soil Ca is from weathered limestone and the weathering of certain primary minerals. This indicates that parent material plays a major role in determining soil Ca concentration. The soil concentration of Ca can vary greatly and is generally related to soil pH. Soils with a pH of 6.0 or greater usually have adequate Ca for plant growth. The majority of plant-available Ca is exchangeable, is much greater than that found in soil solution, and is a divalent cation (Ca^{2+}). An average Ca concentration in soil solution is approximately $0.01\,\mathrm{mol\,l^{-1}}$. The approximate range of exchangeable Ca is 20% to 65% of the cation exchange capacity (CEC) for most soils. Soils may have an average CEC of $15\,\mathrm{cmol_c\,kg^{-1}}$ and 50% Ca saturation. This results in the exchangeable Ca being about 12 000 times greater than soil solution. Some soils contain mineral Ca in the form of precipitated $CaCO_3$ or $CaSO_4 \cdot 2H_2O$ within the soil profile, which results in Ca concentrations much greater than the exchangeable. In the western USA, these soils are usually found in areas of reduced rainfall.

Soil pH and available Ca are usually directly correlated. For example, as pH increases the base saturation increases, thus the amount of exchangeable Ca, Mg, and K increase. These are considered to be basic cations while Al and H are considered acidic cations. Often, liming soils to raise low soil pH will greatly increase Ca concentration in soils. Limestone is the primary liming material and there are different types of limestone depending on the ratio of Ca to Mg in the limestone. Calcite is $CaCO_3$ with essentially no $MgCO_3$ but as the amount of $MgCO_3$ begins to increase within the limestone, the limestones are referred to as calcitic, dolomitic, and dolomite. Dolomite contains about 50% $CaCO_3$ and 50% $MgCO_3$. Other sources of $CaCO_3$ are marl or chalk, which are soft sources of limestone. Choices of liming material may be determined by the nutrient status of Ca or Mg. Another available Ca source is gypsum ($CaSO_4 \cdot 2H_2O$), often used when soil pH is adequate, but additional Ca is required. Industrial by-products that have neutralizing properties such as quick lime, burnt lime, basic slag, fly ash and some sludge materials may also contribute Ca.

Calcium has fewer pathways than other macronutrients and thus, the management of adequate soil Ca is simpler. The loss of Ca is generally by plant uptake, since it does not volatilize and leaching is minimal. For example, Ca is not easily leached because it is held to cation exchange sites more strongly than other nutrients, which is demonstrated by the lyotropic series:

$$\mathrm{Al}^{3+} > \mathrm{Ca}^{2+} > \mathrm{Mg}^{2+} > \mathrm{K}^{+} = \mathrm{NH}_4^{+} > \mathrm{Na}^{+}.$$

The strength of retention increases with the charge on the cations, and also depends on their hydrated size with smaller cations being held more tightly. Inputs into the pool of available Ca must be considered for proper management. Very little available Ca is derived from mineralization of soil organic matter. So the input of available Ca is from weathering of Ca minerals or from supplemental applications of Ca sources as those mentioned earlier. Plant uptake of Ca is another pathway and the uptake characteristic is important in the method of supplying adequate Ca.

The major role of Ca in plants is cell-wall integrity and function, and it plays a role in cell division and elongation. Calcium is bound to negatively charged phosphate groups of membrane lipids. This prevents inorganic and organic molecules from leaking out of the cells. When Ca deficiency continues there is a general disintegration of the membrane and cell-wall leaking. In the whole plant, the meristematic tissue shows the deficiency in the growing point and root tips. Most visual Ca deficiency symptoms occur in fruit and storage tissue, because translocation of Ca is by way of xylem tissue. Thus, low soil solution Ca or very low transpiration rate in humid conditions may cause Ca deficiency. Examples of common symptoms include: apples display ‘bitter pit’ where the surface is pitted with brown spots; tomatoes, peppers, and watermelons have blossom-end rot where there is complete tissue breakdown at the distal end of the fruit; celery has black heart and brussel sprouts have internal browning; and peanuts exhibit ‘pops’ where the hull is completely formed but the kernel did not develop.

Magnesium

The lithosphere contains about 2.1% Mg and soil generally contains 0.5%. Parent material plays a major role in determining available soil Mg concentration. Soil Mg is derived from the weathering of several minerals, much like Ca. Limestones (dolomitic and dolomite) supply Mg in many soils as well as some primary and secondary minerals. Mg, unlike Ca, is contained in several secondary minerals e.g., montmorillonite, vermiculite, illite, and chlorite. Soils rich in these minerals usually have sufficient available Mg and do not need supplements. Much like Ca, most of the available Mg is exchangeable but usually at a lower concentration. Generally Mg soil solution concentration is about 0.001 $\mathrm{mol\,l^{-1}}$, which is lower than soil-solution Ca. The lower availability of Mg can be attributed to the lyotrophic series that indicates Mg can be leached more easily than Ca.

As mentioned earlier with Ca, availability of Mg is also often related to soil pH. Soil Mg is normally increased when applying dolomite or dolomitic limestone when correcting soil pH. When adding Mg without adjusting pH other sources are available; such as Epsom salt ($\mathrm{MgSO_4 \cdot 7H_2O}$), magnesium ammonium phosphate ($\mathrm{MgNH_4PO_4 \cdot H_2O}$), and a more common source Sul-Po-Mag ($\mathrm{K_2SO_4 \cdot 2MgSO_4}$). Other organic sources may supply Mg, but mineralization of organic matter does contribute substantially to Mg availability.

The pathway for Mg is the same as Ca, except Mg is not held to exchange sites as strongly as Ca and can be subject to some leaching sooner than Ca. Uptake and translocation of Mg is different than Ca. Magnesium is very mobile in the phloem and is translocated from older to younger leaves and translocates to storage tissue. The deficiency symptom occurring first in the older tissue, depending on the type, is an intervenal yellowing or chlorosis, in severe cases becoming necrotic. Magnesium is the central component of chlorophyll and acts as a cofactor in most enzymes activating phosphorylation.

Sulfur

Management of soil S is more complicated than most other nutrients and may be considered similar to that of N. Sulfur content of the lithosphere is about 600 ppm and while there is a large range (30–10 000 ppm) in soil S, the average concentration is estimated at 700 ppm. The four major forms of S in soils are: (1) sulfides, (2) sulfates, (3) organic sulfur, and (4) elemental sulfur. Understanding the S cycle in soil is important for proper management of soil S. Sulfur mineralization is very similar to N mineralization with decay products being in a reduced form (sulfides) then oxidized to a sulfate form. Sulfate is subject to reduction, immobilization, fixation by clays, precipitation, anion exchange, leaching and plant uptake. In maintaining adequate soil S fertility all these processes must be considered. The reduction of sulfates to sulfides requires anaerobic conditions in combination with an available C source. Once reduced it may be lost to the atmosphere or precipitated and then be subject to oxidation when conditions become aerobic. The oxidation of sulfides or elemental sulfur causes acidification of soils. This may be desirable when attempting to grow acid-loving plants on initially high pH soils. Major inputs into the pool of available soil sulfur are from soil organic matter, fertilizers, and deposition from the atmosphere (wet or dry). Other minor sources are from irrigation water, plant residues or manures, and sulfur-bearing minerals.

In many cases the available sulfur must be supplemented. The common sources of sulfur are gypsum ($CaSO_4 \cdot 2H_2O$; 19% S), elemental S (flowers of sulfur; 95% S), ammonium sulfate ($(NH_4)_2SO_4$; 24% S) and sulfur coated urea (13–16% S). Sulfur-deficient soils requiring these inputs are usually sandy-textured and the deficiency is usually more prominent in early cool spring months. This occurs mainly for two reasons: mineralization is minimal and earlier mineralized S has been taken up or lost due to previously mentioned pathways. Also, root growth is limited due to cool soil temperature. These reasons often explain why crops may grow out of S deficiencies when soil temperature increases, enhancing mineralization and root growth.

Sulfur is taken up mainly as SO_4, is considered nonmobile and is only translocated in the xylem. Thus, deficiencies usually appear in the younger tissue. Deficiency symptoms are often a general chlorosis in younger leaves. Sulfur is a component of many proteins in plants, but when plants become deficient in S there is an accumulation of non-S proteins and an associated decrease in sugars. The decrease in sugar inhibits chlorophyll production, resulting in a general chlorosis.

List of Technical Nomenclature

°C	Celsius
Al	Aluminum
atm	Atmosphere
Ca	Calcium
$CaCO_3$	Calcium carbonate
$CaSO_4 \cdot 2H_2O$	Gypsum
CEC	Cation exchange capacity
$CO(NH_2)_2$	Urea
$cmol_c\,kg^{-1}$	centimole charge per kilogram
$g\,kg^{-1}$	gram per kilogram
H	Hydrogen
HPO_4	Monohydrogen phosphate
H_2PO_4	Dihydrogen phosphate
K	Potassium
KCl	Potassium chloride
$K_2SO4 \cdot 2MgSO_4$	Sul-Po-Mag
$kg\,ha^{-1}\,yr^{-1}$	kilogram per hectare per year
Mg	Magnesium
$MgCO_3$	Magnesium carbonate
$MgNH_4PO_4 \cdot H_2O$	Magnesium ammonium phosphate
$MgSO_4 \cdot 7H_2O$	Epsom salt
$mg\,kg^{-1}$	milligram per kilogram
$mg\,l^{-1}$	milligram per liter
$mol\,l^{-1}$	moles per liter
N	Nitrogen
N_2	Dinitrogen
Na	Sodium
NH_3	Ammonia
NH_4^+	Ammonium
$(NH_4)_2SO_4$	Ammonium sulfate
NO_2	Nitrite
NO_3	Nitrate
N_2O	Nitrous oxide
O_3	Ozone
P	Phosphorus
ppm	part per million
S	Sulfur
SO_4^{2-}	Sulfate

See also: **Eutrophication**; **Nitrogen in Soils:** Plant Uptake; **Organic Residues, Decomposition**; **Phosphorus in Soils:** Biological Interactions; **Pollution:** Groundwater; **Sulfur in Soils:** Biological Transformations

Further Reading

Brady NC and Weil RR (2002) The nature and properties of soils, 13th edn. Upper Saddle River, NJ: Prentice Hall.

Havlin JL, Tisdale SL, Nelson WL, and Beaton JD (1999) *Soil Fertility and Fertilizers: An Introduction to Nutrient Management*, 6th edn. Upper Saddle River, NJ: Prentice Hall.

Hewitt EJ (1984) The effects of mineral deficiencies and excesses on growth and composition. In: Bould C, Hewitt EJ, Needham P, and Robinson JBD (eds) *Diagnosis of Mineral Disorders in Plants,* vol. 1, pp. 54–110. New York, NY: Chemical Publications.

Lindsay WL (1979) *Chemical Equilibria in Soils*. New York, NY: Wiley-Interscience Publications.

Mengel K and Kirkby EA (1987) *Principles of Plant Nutrition*, 4th edn. Bern, Switzerland: International Potash Institute.

Pierzinsky GM, Sims JT, and Vance GF (2000) *Soils and Environmental Quality*, 2nd edn. Boca Raton, FL: CRC Press.

MACROPORES AND MACROPORE FLOW, KINEMATIC WAVE APPROACH

P F Germann, University of Bern, Bern, Switzerland

Introduction

Discussion about the significance of structural soil voids carrying water flow dates back at least to the mid nineteenth century when Dupuit applied Darcy's law to two-dimensional flow towards drainage ditches. Many soil scientists at that time were convinced that water flows almost exclusively through wide structural soil pores, which are thought of as any cavity wider than expected from the mere arrangement of soil particles. Structural voids are due to processes like cryo- and bioturbation, stress and strain, shrinking and swelling, and internal transport of particulate and dissolved matter. They are called macropores and permit small water amounts to move quickly through soils.

The dispute intensified and subsided at times depending on the development of various approaches to water flow and transport of solutes and particles in water-saturated and nonsaturated soils. However, the Richards equation has evolved to become the predominant approach to calculating water flow in unsaturated soils since the 1930s, and solute transport has been treated as a convection–dispersion process since the 1960s. Darcy's law, Richards equation and the convection–dispersion approaches to flow and transport in unsaturated soils do not require information on soil structure. Therefore, any new approach to flow and transport in soils that requires structural information or that is designed to explain the effects of soil structure on transport tends to rekindle the controversy and perplexity.

The last three decades have seen intense development in field and laboratory instrumentation, electronic data acquisition and processing, and a generally increased demand for information about environmental processes. Thus, it was recognized that information on soil structure is required to explain better the observed transport of water, solutes, and particles in soils. Knowledge about the operational links between structure and transport processes in soils is still in its infancy despite the demand for suitable scientific and applied approaches. Numerous computer codes simulate the transport processes, although no commonly accepted theoretical framework of rapid flow and transport in structured soils has evolved so far.

Macropores are thought to be structural voids that are responsible for the observed deviations of transport from predictions with the Richards equation and the convection–dispersion approaches. Many empirical measures about the minimum widths and lengths of macropores have been proposed. Because of the nonexistence of alternatives to the Richards equation and the convection–dispersion approaches, there are, at present, few ways of functionally defining the dimensions of macropores with respect to transport processes.

These difficulties result in two consequences for this article: (1) terms like 'macropores' refer to arbitrarily defined soil structural voids. 'Macropore flow' and 'transport' infer processes strongly related to them, but without being properly based on independent and more general flow theory. Therefore, the process-oriented expressions 'preferential flow' and 'preferred' flow paths will be used; (2) because of the many evolving alternate approaches to transport in structured soils, this article cannot assume to be complete and final. Thus, reference to past achievements may illustrate the dynamics of the topic and hint at its evolution.

Early Considerations on Flow and Transport in Soils

Poiseuille, presumably the first researcher rigorously dealing with fluid flow in closed conduits, stated in

1844 that laminar flow through a cylindrical pipe is proportional to the pressure gradient and the fourth power of the pipe's diameter, and inversely proportional to the length of the pipe and the viscosity of the fluid.

Darcy, possibly inspired by the emerging laws of resistance-to-flow – Ohm's law dealing with electric currents, Poiseuille's law dealing with flow of liquids in closed conduits, Laplace's and Fourier's work on heat flow – found in 1856 that flow through a completely water-saturated porous medium is proportional to the medium's cross-sectional area and the gradient of the hydraulic head, and inversely proportional to the length of the macroscopic flow path. The factor of proportionality is called hydraulic conductivity of the water-saturated medium.

In the 1880s agricultural scientists in England reported an early breakthrough concerning drainage flow of water with a chemical composition similar to that of rain water, whereas the composition of drainage flow usually reflected the chemical soil environment. This phenomenon was later dubbed as 'by-pass flow'.

At the beginning of the twentieth century, agricultural plant physiologists discovered that the air-filled soil pores must exceed 5% by volume to guarantee proper aeration of pot-grown sunflowers. At about the same time, Swiss forest scientists, in their quest to justify vast reforestation projects by demonstrating the forests' superiority over many other types of canopies to mitigate peak runoff in steep upland catchments, developed a test that linked infiltration with soil porosity information. A steel cylinder 100 mm high and with a volume of 1 l was completely driven into the soil and covered with a wire mesh. A second cylinder of the same dimensions was tightly set on the first one and filled with water to its rim, and the time for the 100 mm of water to infiltrate completely was recorded. Soil cores were sampled nearby with the same cylinder type. They were completely saturated with water, weighed, drained on a gravel bed for 24 h and weighed again. The weight difference was divided by the density of water and the volume of the core, and its multiplication with 100% resulted in the core's air capacity. More than 50 pairs of infiltration times vs air capacities led to the results shown Figure 1.

The results support the 5%-threshold reported by the plant physiologists. In the 1970s, additional support of the 5%-threshold followed from measurements of oxygen conductivity vs air-filled porosity. The experiments indicate that pores quickly draining to about 5% by volume are hydraulically well-connected.

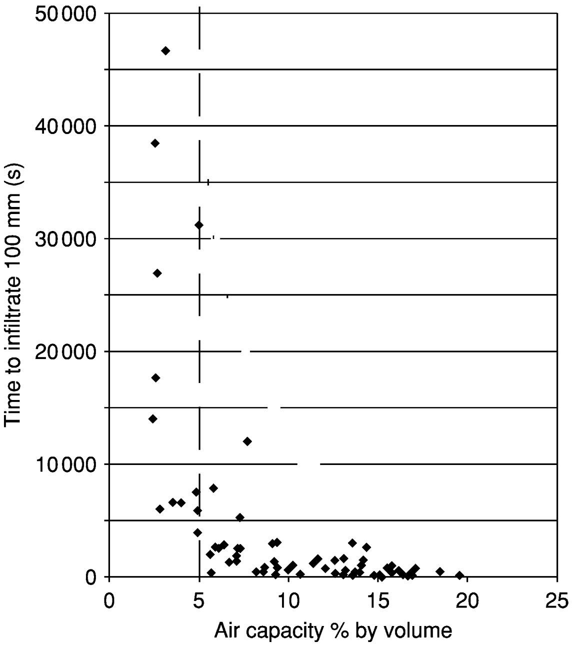


Figure 1 Time to infiltrate 100 mm of water vs air capacity. The empirical threshold at 5% by volume approximately separates the easily-drainable soil samples from those with reduced permeability.

Deviations from Richards Equation and Convection–Dispersion Approaches

Some of the early approaches to flow in unsaturated soils included information on porosity; however, many were not theoretically based and they had to yield to a more rigorous approach. In the late 1920s, Richards included the capillary potential Ψ of the water in unsaturated soils in the hydraulic gradient. He then combined Darcy's law with the volume balance and assumed instantaneous equilibration during transient flow between water content θ and Ψ according to the retention curve. The assumptions led to a quasi-hydrostatic approach to transient flow that expresses the diffusion of water potentials. He called the hydraulic conductivity function $K(\Psi \text{ or } \theta)$, an expression of proportionality.

Veihmeier's empirical concept of field capacity is close to the one of the 5% threshold of air-filled pore space. It hints at domains not predictable by the Richards equation. However, no theoretical explanation followed, presumably due to the lack of alternative approaches to the Richards equation.

In the 1960s, the concept of hydrodynamic dispersion was introduced to explain the spread of concentrations during solute transport through porous media. Ideally, half of the concentration difference is attained in constant drainage flow when one pore volume of soil water has been replaced. Likewise, at some depth in the soil the normalized time variation of the solute concentration follows the S-shaped integral of the normal distribution function.

The increasing number of *in-situ* experiments and observations of flow at the scales of soil profiles, the vadose zone to unconfined groundwater tables, hillslopes, and first-order catchments, increasingly demonstrated that small-volume fractions of soil moisture may move rapidly and over long distances, and that the processes were not satisfactorily explained by the Richards equation. The deviating observations are summarized under the term of non-equilibrium conditions with respect to the Richards equation, and macropores of some length are thought to be the cause. In the late 1970s, the hunt started for concepts that supposedly link structure with transport.

Particular interpretations of water retention curves, Ψ vs θ, in terms of the equivalent pore diameters near saturation, were thought to transform nonequilibrium flow conditions to conform to the Richards equation. Also dual-porosity media have been proposed, with one fraction responding rapidly to infiltration and the other fraction showing more gradual responses, allowing for water to flow from one fraction to the other depending on the hydraulic gradient between them. In addition, observed breakthroughs of dye and particle tracers were frequently faster and distorted when compared with the predictions of the convection–dispersion approaches. A broad variety of transport concepts are offered in the literature based on dual-porosity, neural networks, fuzzy logic, and fractal dimensions.

However, the explanation of finger flow is the most subtle case of nonequilibrium flow and does not require structural pores. The hysteresis in the Ψ-θ relationship initiates finger flow that shows in heterogeneous water breakthroughs in soils in the shapes of vertical fingers. Various theoretical approaches to finger flow have been developed, all of which indicate intricate relations between soil and flow parameters.

So far, macropores and macropore flow, preferential flow and preferred flow paths as well as nonequilibrium flow and transport are explicitly or implicitly defined in contrast to ‘ordinarily’ sized pores and flows at equilibrium with respect to the Richards equation and the convection–dispersion concepts.

Flow paths are frequently visualized *in situ* with dye tracers that are applied to the infiltrating water with subsequent preparation and recording of the stained areas. The staining procedures are applied to sample sizes ranging from typical thin sections covering a few square centimeters to pedons and polypedons. The color patterns provide a visualization of infiltration and indicate the paths taken by the water; however, they do not provide information on the flow dynamics and reveal neither the quantity nor the duration of flow along them. Thus, they are of limited value when developing quantitative approaches to flow and transport in structured soils.

Alternative Approaches to Flow and Transport

General Requirements

Alternative approaches to the Richards equation dealing with transient water flow in unsaturated structured soils have to go back to the basic Navier–Stokes equations to avoid the assumption of equilibrium along the Ψ-θ relationship.

Flow and transport in structured soils and in their artificial physical counterparts are increasingly visualized with computer-aided tomography using X-ray, nuclear magnetic resonance (NMR), or various other radiation techniques. Information on the three-dimensional structure and transport is then used to solve the appropriate Navier–Stokes equations at the pore scale by either applying Boltzman transformations or lattice gas simulations. These process-oriented studies are laborious and require a combination of sophisticated tools and skills. They provide excellent insight into the processes, but up-scaling their results to larger soil volumes is still challenging, mainly because of the pronounced nonlinearity of the processes involved.

According to Newton, a solid body is accelerated when a constant force is acting on it, whereas the same constant force acting on a fluid results in its moving with a constant velocity after a short period of acceleration. Likewise, a solid body gains momentum, whereas a fluid achieves constant momentum when both are exposed to the same constant force. The difference is due to the fluid’s viscosity which dissipates linear momentum $\rho \cdot \partial v/\partial f$ perpendicular to the direction of flow towards the solid. However, Richards equation applies to flow in soils as long as momentum completely dissipates within a representative elementary volume, i.e., when the Richards condition is fulfilled.

Preferential flow may expressly consider linear momentum and kinetic energy of transient flow in structured soils. Dissipation of linear momentum during laminar flow results in a shear force in the direction opposite to the flow driving force. Newton’s law of shear states that:

$$\tau(f) = -f \cdot g \cdot \rho = \eta \cdot \rho \cdot \frac{\partial v}{\partial f} \qquad [1]$$

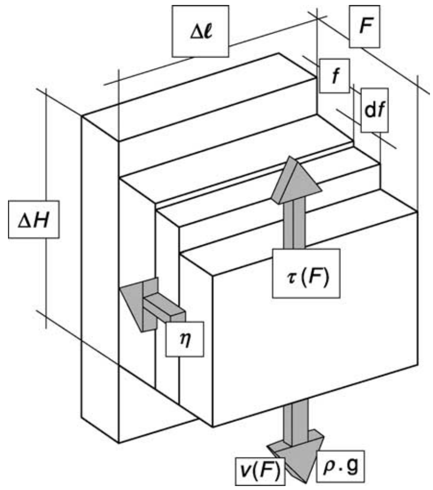


Figure 2 Schematic representation of Newton's law of shear. F(m) Film thickness with its variable f; g (m s^{-2}) acceleration due to gravity; ρ (kg m^{-3}) density of water; $\Delta\ell$ (m) contact length of the water film with the parts at rest; ΔH (m) macroscopic length of coherent water films; η (m^2 s^{-1}) kinematic viscosity dissipating momentum towards the resting part of the soil-water system; $\tau(F)$ (Pa) and $v(F)$ (m s^{-1}) are the shear force and the velocity of the lamina at F.

where $\tau(f)$ (Pa) is the shear force as a function of the film thickness f(m) acting in the opposite direction of gravity (gravity is assumed to be the only flow driving force), g ($=9.81\,\mathrm{m\,s^{-2}}$) is acceleration due to gravity pointing vertically downwards, ρ ($=1000\,\mathrm{kg\,m^{-3}}$) is the density of water, η (approx. $10^{-6}\,\mathrm{m^2\,s^{-1}}$) is the kinematic viscosity of water (i.e., the coefficient of momentum dissipation), and $v(f)$ (m s^{-1}) is the velocity of a water lamina at distance f from the parts at rest. Figure 2 illustrates the definitions of the parameters.

Integration of eqn [1] leads to the velocity profile:

$$v(f) = \frac{1}{2\eta} \cdot f^2 \cdot g \qquad [2]$$

Mass and momentum of a layer at f are:

$$\mathrm{d}M = \Delta H \cdot \Delta\ell \cdot \mathrm{d}f \cdot \rho \qquad [3]$$

and:

$$\mathrm{d}I(f) = \mathrm{d}M \cdot v(f) = \frac{1}{2\eta}\Delta\ell \cdot g \cdot \rho \cdot \Delta H \cdot f^2 \cdot \mathrm{d}f\,(\mathrm{kg\,m^{-1}\,s^{-1}}) \qquad [4]$$

where ΔH is the macroscopic length of well-connected water films. Integration of eqn [4] within $0 \leq f \leq F$ yields horizontal momentum dissipation due to ρ, g and η from the mobile liquid towards the resting parts, like the solid matrix or stagnant water:

$$I_\mathrm{h}(F) = -\frac{1}{6\eta}F^3 \cdot \rho \cdot g \cdot \Delta H \cdot \Delta\ell\,(\mathrm{kg\,m^{-1}\,s^{-1}}) \qquad [5]$$

The combination of eqns [1] and [2] yields the vertical momentum of the water film:

$$I_\mathrm{v}(F) = \frac{1}{6\eta}F^3 \cdot \rho \cdot g \cdot \Delta H \cdot \Delta\ell\,(\mathrm{kg\,m^{-1}\,s^{-1}}) \qquad [6]$$

Steady flow occurs when $I_\mathrm{v} = I_\mathrm{h}$, acceleration when $I_\mathrm{v} > I_\mathrm{h}$ and deceleration when $I_\mathrm{v} < I_\mathrm{h}$.

Mobile volumetric soil moisture is:

$$w = \frac{\Delta H \cdot F \cdot \ell}{\Delta H \cdot A}(\mathrm{m^3\,m^{-3}}) \qquad [7]$$

where A (m^2) is the cross-sectional area of the soil and $\ell = \Sigma\Delta\ell$(m) is the total contact length of w in A. Integration of eqn [2] under the consideration of eqn [7] leads to the volume flux density of momentum-dominated flow as:

$$q = b \cdot w^a (\mathrm{m\,s^{-1}}) \qquad [8]$$

where the conductance is:

$$b = \frac{g}{2 \cdot a \cdot \eta} \cdot \frac{A^2}{\ell \cdot \Delta H}(\mathrm{m\,s^{-1}}) \qquad [9]$$

The dimensionless exponent a is related to the channel geometry. Flow in a cylinder, as in Hagen-Poiseuille's law, results in $a = 2$, whereas flow along a plane or in a planar fissure, also referred to as the cubic law, results in $a = 3$.

Kinematic Wave Theory

The application of eqns [1] to [9] to transient laminar flow leads to the kinematic wave theory as developed by Lighthill and Witham in the early 1950s. It combines Eulerian and Lagrangean concepts of flow where the celerity c (m s^{-1}) describes the velocity of a particular wave property.

The water balance of a kinematic wave is:

$$\frac{\partial q}{\partial t} + c \cdot \frac{\partial q}{\partial z} = 0 \qquad [10]$$

where t and z are time and depth coordinates, the latter is positive vertically down.

The initial and boundary conditions of a rectangular input pulse to the soil surface of duration t_s and volume flux density q_s are:

$$0 \leq z \leq \infty \quad t \leq 0 \quad w(z,t) = 0 \quad q(z,t) = 0 \qquad [11]$$

$$z = 0 \quad t \leq 0 \quad \text{and} \quad t \geq t_\mathrm{S} \quad w(z,t) = 0 \quad q(z,t) = 0 \qquad [12]$$

$$0 \leq t \leq t_\mathrm{S} \quad w(z,t) = w_\mathrm{S}; \quad q(z,t) = q_\mathrm{S} = b\,.\,w_\mathrm{S}^a \qquad [13]$$

A sharp wetting shock front starts moving vertically down at $t = 0$ with a celerity of:

$$c_W = \frac{q_S}{w_S} = b \cdot w_S^{a-1}\ (\mathrm{m\,s^{-1}}) \qquad [14]$$

The position of the wetting shock front as function of time is:

$$z_W(t) = t \cdot b \cdot w_S^{(a-1)}\ (\mathrm{m}) \qquad [15]$$

and the time lapsed for the wetting front to move to depth z is:

$$t_W(z) = \frac{z}{b \cdot w_S^{(a-1)}} \qquad [16]$$

At $t = t_S$, water input and mobile soil moisture at the soil surface drop to zero. However, the sharp draining shock front at the surface decays to an infinite number of layers (i.e., laminae) in the range of $0 \le w \le w_S$. Each lamina moves with a celerity:

$$c(w) = \frac{\mathrm{d}q}{\mathrm{d}w} = a \cdot b \cdot w^{(a-1)} = \frac{z(w)}{(t - t_S)}\ (\mathrm{m\,s^{-1}}) \qquad [17]$$

corresponding to the position f within the water film. The fastest lamina is the draining shock front whose celerity is:

$$c_D = \frac{\mathrm{d}q_S}{\mathrm{d}w_S} = a \cdot b \cdot w_S^{(a-1)}\ (\mathrm{m\,s^{-1}}) \qquad [18]$$

The depth of the draining front as function of time is:

$$z_D(t) = (t - t_S) \cdot a \cdot b \cdot w_S^{(a-1)} (\mathrm{m}) \qquad [19]$$

and the time lapsed to its arrival at z is:

$$t_D(z) = t_S + \frac{z}{a \cdot b \cdot w_S^{(a-1)}} \qquad [20]$$

The position in depth and time of any mobile water content in the range of $0 \le w \le w_S$ and at time $t \ge t_S$ follows from solving eqn [17] for:

$$w(z,t) = \left[\frac{z}{(t - t_S) \cdot a \cdot b}\right]^{\frac{1}{(a-1)}} \qquad [21]$$

Because $c_D = a \cdot c_w$ the draining front intercepts the wetting front at time:

$$t_I = t_S \frac{a}{(a-1)}\ (\mathrm{s}) \qquad [22]$$

and depth:

$$z_I = t_S \cdot \frac{a}{(a-1)} \cdot b \cdot w_S^{(a-1)}\ (\mathrm{m}) \qquad [23]$$

The total water volume that has infiltrated into the soil and that is contained in the kinematic wave is:

$$V = t_S \cdot q_S = t_S \cdot b \cdot w_S^{a} (\mathrm{m}) \qquad [24]$$

Integration of eqn [21] i.e., $\int_0^{zF(t)} w(z,t)\mathrm{d}_z = V$, yields the position of the wetting shock front for times $t \ge t_I$ and depths $z_F(t) \ge z_I$:

$$z_F(t) = a \cdot b \cdot w_S^{(a-1)} \cdot \left[\frac{t_S}{(a-1)}\right]^{\frac{(a-1)}{a}} \cdot (t - t_S)^{1/a} \qquad [25]$$

Inserting eqn [25] into eqn [21] yields mobile soil moisture at the wetting front at $z_F(t)$:

$$w_F(t) = w_S \cdot \left[\frac{t_S}{(a-1) \cdot (t - t_S)}\right]^{1/a} \qquad [26]$$

and the celerity of the wetting front is:

$$c_F(t) = b \cdot w_S^{(a-1)} \cdot \left[\frac{t_S}{(a-1) \cdot (t - t_S)}\right]^{\frac{(a-1)}{a}} \qquad [27]$$

Equation [21] applies to the ranges of time $t > t_D$ and depth $0 < z < z_D$, z_F. It describes the trailing wave of a kinematic wave. The combination of eqn [19] with eqn [21] yields the decline of mobile soil moisture at a particular depth after the passing of the draining front ($t_S \le t \le t_I$) or the wetting front ($t \ge t_I$):

$$w(z_D, t) = w_S \cdot \left[\frac{(t_D - t_S)}{(t - t_S)}\right]^{\frac{1}{(a-1)}} \quad \text{or}$$

$$w(z_F, t) = w_S \cdot \left[\frac{(t_F - t_S)}{(t - t_S)}\right]^{\frac{1}{(a-1)}} \qquad [28]$$

A complete kinematic wave is shown in **Figure 3**. Application of eqn [8] leads to the volume flux densities, and a combination of eqns [7] and [9] provides the film thickness and the contact length of mobile soil moisture with the resting parts.

Kinetic energy of flow per unit volume of mobile soil water is:

$$E_{\mathrm{kin,w}} = \rho \cdot \frac{q^2}{2 \cdot w^2} = \rho \cdot b^2 \cdot w^{2(a-1)}\ (\mathrm{Pa}) \qquad [29]$$

and momentum of flow per unit volume of mobile soil water is:

$$i_w = \rho \cdot \frac{q}{w} = \rho \cdot \mathrm{b} \cdot \mathrm{w}^{(a-1)} (\mathrm{kg\,m^{-2}\,s^{-1}}) \qquad [30]$$

Application of the Kinematic Wave Model

Sprinkling at the soil surface during $0 < t < t_S$ with an intensity of q_S provides for the boundary conditions, eqns [11] to [13]. Because t_I depends only on

t_S, eqn [22], the duration of the latter is selected long enough for the maximum depth of observation to be within $0 < z < z_I$. The kinematic wave approach is applied either to drainage flow at some depth or to soil moisture variations at several depths. Temporal soil moisture variations are typically recorded with time-domain reflectometry (TDR) measurements using horizontally installed waveguides and frequencies of data recording higher than 1/300 Hz. (The system takes repeated water-content readings at each waveguide at intervals of 300 s, i.e., every 5 min.) A time series of drainage flow or soil moisture in the depth range of $0 < z < z_I$ follows three stages:

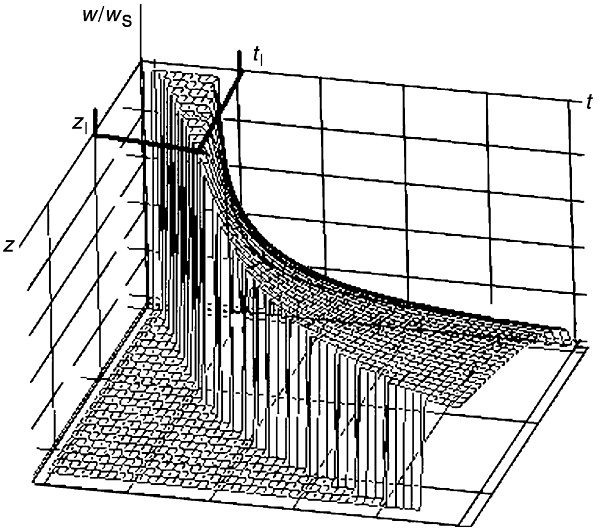


Figure 3 Schematic representation of a kinematic wave normalized with respect to the wave's amplitude w_S.

$$\text{Stage 1: } 0 < 1 < t_W(z) \quad w = 0 \quad q = 0 \qquad [31]$$

$$\text{Stage 2: } t_W(z) < t < t_D(z) \quad w = w_S \quad q = q_S \qquad [32]$$

$$\text{Stage 3: } t_D(z) < t < \infty \quad w = w_S \cdot \left(\frac{t_D(z) - t_S}{t - t_S}\right)^{1/(a-1)}$$

$$q = q_S \cdot \left(\frac{t_D(z) - t_S}{t - t_S}\right)^{a/(a-1)} \qquad [33]$$

An example of a model application to data is shown in Figure 4, where the overall soil moisture maximum w_S is replaced by its local amplitude w_{max}.

Soil layers of lower hydraulic conductivity, mainly when they are lacking paths of preferential flow, may backlog water with subsequent distortion of the wave's shape. Water backlogging may also temporarily produce positive pressures, thus requiring adaptation of eqn [1]. In addition, capillary attraction may adsorb parts of the rapidly flowing water at low antecedent degrees of saturation, thus markedly reducing the effects of preferential flow at depths. These more local effects are thought to distort a kinematic wave such that no single set of parameters *a* and *b* has so far been found capable of quantifying flow over a considerable depth range. Thus, the kinematic wave model needs calibration to data at the various depths of measurement. However, once calibrated, the model allows for the estimation of flow parameters at particular depths, as their approximate ranges are compiled in Table 1.

Moreover, the local quantification of flow with the kinematic wave model at various depths provides information on the water sorption processes during

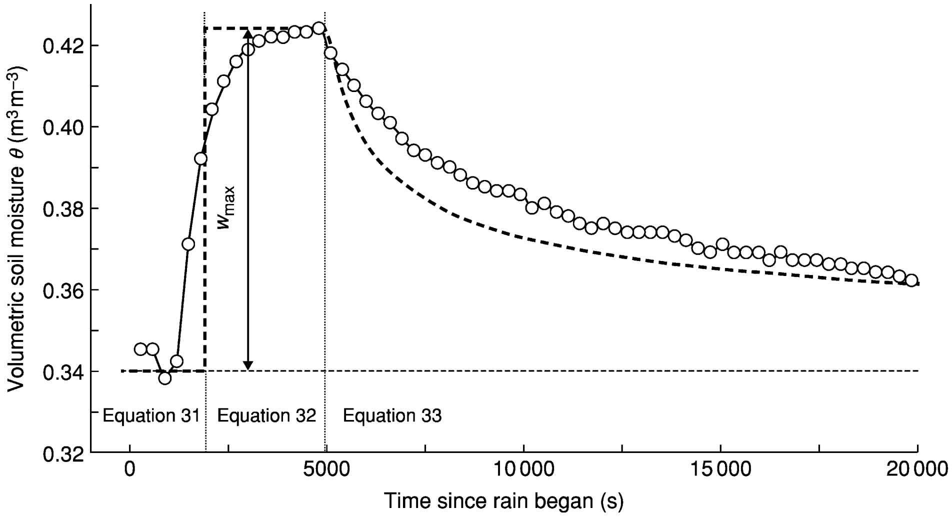


Figure 4 Soil moisture variation at depth 0.35 m due to sprinkling 125 mm during 4500 s (run 1, site Rütti). The local amplitude w_{max} replaces w_S; t_W and t_D are the arrival times of the wetting and draining fronts. The ranges of equation applicability are also shown. (See also Figure 5.) --- model; -o- data.

Table 1 Ranges of flow parameters

Parameter	*Symbol*	*Unit*	*Minimum*	*Typical value*	*Maximum*
Input rate	q_s	$\mathrm{m\,s^{-1}}$	3.10^{-5}	8.10^{-6}	2.10^{-6}
Mobile soil moisture	w_S	$\mathrm{m^3\,m^{-3}}$	0.003	0.03	0.12
Exponent	a	—	2.0	3.0	6.0
Conductance	b	$\mathrm{m\,s^{-1}}$	0.001	0.3	5.0
Contact length of w_S per cross-sectional area	ℓ	$\mathrm{km\,m^{-2}}$	0.1	5.0	30
Film thickness	F	μm	2.0	15	50
Momentum per w_S	i_w	$\mathrm{kg\,m^{-2}\,s^{-1}}$	0.01	0.5	1.0
Kinetic energy per w_S	$E_{\mathrm{kin,w}}$	Pa	10^{-5}	10^{-4}	10^{-3}

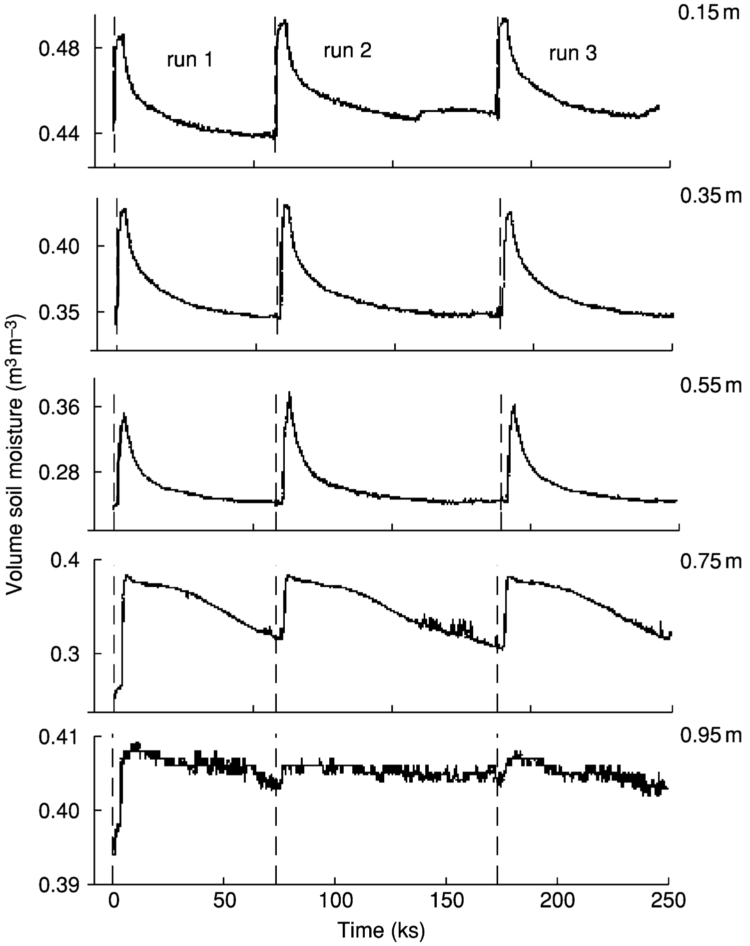


Figure 5 Soil moisture variation at depths 0.15, 0.35, 0.55, 0.75, and 0.95 m due to sprinkling 125 mm three times during 4500 s (site Rütti). The kinematic waves are easily discernible at the upper three depths of measurements. The convex reactions at depth 0.75 m indicate water backlogging above an impermeable layer between this and the 0.95-m depth. Note the different scales on the ordinates.

preferential flow between the two particular depths of soil moisture measurements. Examples of θ (Z, t), variations indicating deviations from the kinematic wave approach are shown in Figure 5.

The contact lengths and film thickness indicate that preferential flow is through a large number of small pores and not exclusively through 'macropores' of not-so-well-defined geometries. The range of the amplitudes includes the empirical threshold of 5% by volume of air capacity and air-filled porosity, as shown in Figure 1. On one hand, the kinetic energy per unit of flowing water is orders of magnitude smaller than the usual range of capillary potential in unsaturated soils, indicating their mutual independence. On the other hand, momentum per unit of flowing soil water is about 100 times greater than its counterpart of the infiltration rate, indicating channeling of flow in soils. Regarding the fact that capillarity was observed in a pipe of 12 mm inner diameter, preferential flow may concur with Richards-type flow in an approximate pore size range of $6\,\text{mm} > R > 30\,\mu\text{m}$, which represents a relative range of about 1:200. Momentum of infiltration in combination with a minimum degree of antecedent saturation seem to be the two decisive flow parameters that may trigger preferential flow if the geometry of the soil structural voids permit a wave's coherence over a considerable length of flow paths.

Soil hydrologists deal mainly with vertical preferential flow, whereas catchment hydrologists relate preferential flow with lateral subsurface storm flow. The link between the two aspects requires bending of flow from the vertical to the lateral direction at some soil depth. Research on momentum of flow includes the investigation of its scalar and its vector. Early experiments in a layered soil on a hillslope using three-dimensional arrangements of TDR-waveguides indeed revealed bending of preferential flow.

Further Developments of Kinematic Wave Approaches to Preferential Flow

The kinematic wave approach, eqns [1] to [30], explains local soil moisture variations reasonably well. However, no single wave has been found so far that would explain infiltration at the profile scale. Water losses from the preferred flow paths into the soil matrix were assumed to be responsible for the deviations. Therefore, a sink function was introduced in the mid-1980s that rides on the kinematic wave. The water balance of a kinematic wave with a sink-function extension is:

$$\frac{\partial q}{\partial t} + c \cdot \frac{\partial q}{\partial z} = c \cdot w \cdot r \qquad [34]$$

with the sink strength $r\,(\text{s}^{-1})$ defined as:

$$\frac{\partial w}{\partial t} = r \cdot w\,(t) \qquad [35]$$

Due to the sink the wetting front may not penetrate beyond the depth of:

$$z^* = \frac{a \cdot b \cdot w_{\text{S}}^{(\text{a}-1)}}{r \cdot (a-1)} \qquad [36]$$

However, no improvement was found when calibrating the extended kinematic wave model, eqns [34] to [36], against data. It seems that water losses from the kinematic wave to the surrounding matrix pores are more complicated than assumed with eqn [35].

An approach based on a traveling diffusive wave was introduced by Di Pietro in 2002. Its water balance is:

$$\frac{\partial q}{\partial t} + c \cdot \frac{\partial q}{\partial z} = u\frac{\partial^2 q}{\partial z^2} \qquad [37]$$

where u ($\text{m}^2\,\text{s}^{-1}$) treats the wave's spreading during flow. The approach based on eqn [37] was successfully calibrated against observed drainage flow from a macroporous soil.

Applications

The concepts of preferential flow may apply to areas outside the hydrology of soils and hillslopes.

The transport of particulate matter through soils is related to the kinetic energy of flow, eqn [29], which results in a drag force acting on suspended particles. Recent investigations demonstrated that breakthrough of latex beads with diameters ranging from 0.5 to 1.75 μm and having electrically neutral surface charges was more strongly related with the kinetic energy of flow than with the variable particle diameters. In contrast, breakthrough of bacteriophages showed very strongly the effects of their variably charged surfaces.

In many areas of the industrialized world, soil compaction is the result of the employment of increasingly more efficient and thus increasingly heavier agricultural machinery. Rotating wheels may not only compact a soil in the vertical direction but knead it back and forth to some depth. The pores allowing preferential flow, rapid drainage, and efficient aeration may be squeezed at times below the plow layer. Soil strength decreases with increasing soil moisture. Thus a vicious cycle starts in that some squeezed preferred flow paths delay drainage and prolong the periods of reduced soil strength. Compaction potential thus increases if the successive soil operation happens under relatively wet soil conditions. Infiltration experiments with subsequent soil moisture recording with TDR-equipment at several soil depths allows an assessment of whether the structural pores are still adequate for drainage and aeration in the 5%-volume fraction range shown in **Figure 1**.

Summary

Poiseuille's early application of momentum dissipation to flow in closed conduits was not considered suitable to deal with flow in porous media. Darcy's law and the Richards equation were developed to deal with such flow. They do not require *a priori* information on soil structure, and the latter is based on the diffusion of potential.

Momentum dissipation approaches were applied to laminar flow in structured and nonsaturated soils with some success, still leaving open many areas for research. The major breakthrough in understanding flow in nonsaturated and structured porous media is expected to come from diffusion wave approaches that seem to lead from strict momentum dissipation and to include diffusive processes. Solute and particle transport in structured and nonsaturated soils will possibly follow similar tracks.

The relative range of pore diameters that are able to carry Richards-type flow as well as preferential flow is about 1:200. Momentum of infiltration is thought to trigger preferential flow when it is exceeding some minimum. The importance of the uppermost soil layer as being able either to transmit or mitigate momentum to deeper layers for the initiation of or prevention of preferential flow is also implicated.

List of Technical Nomenclature

η	Kinematic viscosity of water ($m^2 s^{-1}$)
θ	Volumetric soil moisture ($m^3 m^{-3}$)
ρ	Density of water ($kg\,m^{-3}$)
τ	Shear force per cross-sectional area (Pa)
ψ	Capillary potential (Pa)
A	Cross-sectional area of soil (m^2)
a	Exponent of kinematic wave (–)
b	Conductance of kinematic wave ($m\,s^{-1}$)
c	Celerity of kinematic wave ($m\,s^{-1}$)
c_D	Celerity of the draining front ($m\,s^{-1}$)
c_F	Celerity of the wetting front after interception ($m\,s^{-1}$)
c_W	Celerity of the wetting front ($m\,s^{-1}$)
$E_{kin,w}$	Kinetic energy per unit volume of mobile water (Pa)
F	Thickness of water film (m)
f	Film thickness variable (m)
g	Acceleration due to gravity ($m\,s^{-2}$)
H	Vertical macroscopic length of connected water films (m)
I	Momentum of water film ($kg\,m\,s^{-1}$)
I_h	Horizontal momentum loss of water film ($kg\,m\,s^{-1}$)
I_v	Vertical momentum of water film ($kg\,m\,s^{-1}$)
i_W	Momentum per unit volume of mobile water ($kg\,m^{-2}\,s^{-1}$)
K	Hydraulic conductivity ($m\,s^{-1}$)
ℓ	Contact length of mobile water (m)
M	Mass of soil water (kg)
q	Volume flux density ($m\,s^{-1}$)
q_s	Volume flux density of input ($m\,s^{-1}$)
R	Radius (m)
r	Sink strength ($m\,s^{-1}$)
t	Time (s)
t_D	Arrival time of the draining front (s)
t_I	Time of front interception (s)
t_S	End of input to the soil surface (s)
t_W	Arrival time of the wetting front (s)
u	Coefficient of wave diffusion ($m^2 s^{-1}$)
V	Water volume of kinematic wave per cross-sectional area (m)
v	Velocity of water film ($m\,s^{-1}$)
w	Mobile soil moisture ($m^3 m^{-3}$)
w_F	Mobile soil moisture at the front after front interception ($m^3 m^{-3}$)
w_{max}	Local amplitude of the kinematic wave ($m^3 m^{-3}$)
w_S	Mobile soil moisture at the surface during infiltration ($m^3 m^{-3}$)
Z	Specified soil depth (m)
z	Depth vertically down (m)
z^*	Maximum depth of front penetration (m)
z_D	Depth of draining front (m)
z_F	Depth of wetting front after front interception (m)
z_I	Depth of front interception (m)
z_W	Depth of wetting front (m)

See also: **Aggregation:** Physical Aspects; **Fauna**; **Morphology**; **Structure**; **Subsoiling**; **Unstable Flow**; **Vadose Zone:** Microbial Ecology

Further Reading

Bosh D and King K (eds) (2001) *Preferential Flow: Water Movement and Chemical Transport in the Environment – Proceedings of the 2nd International Symposium 3–5*

January 2001. St Joseph, Michigan: American Society of Agricultural Engineers.

Di Pietro L, Ruy S, and Capowiez Y (2003) Predicting preferential water flow in soils by travelling diffusive waves. *Journal of Hydrology* 278: 64–75.

Germann PF (2001) A hydromechanical approach to preferential flow. In: Anderson MG and Bates PD (eds) *Model Validation – Perspectives in Hydrological Sciences*, pp. 233–260. Chichester: John Wiley.

Germann PF and Di Pietro L (1999) Scales and dimensions of momentum dissipation during preferential flow in soils. *Water Resources Research* 35(5): 1443–1454.

Gish TJ and Shirmohammadi A (eds) (1991) *Preferential Flow – Proceedings of the National Symposium 16–17 December 1991*. St Joseph, Michigan: American Society of Agricultural Engineers.

Selim HM and Ma L (eds) (1998) *Physical Nonequilibrium in Soils: Modelling and Application*. Ann Arbor: Ann Arbor Press.

Selim HM and Sparks DL (eds) (2001) *Physical and Chemical Processes of Water and Solute Transport/Retention in Soil*. Soil Science Society of America Special Publication No. 56. Madison: Soil Science Society of America.

Magnesium in Soils *See* **Calcium and Magnesium in Soils**

MANURE MANAGEMENT

J T Sims, University of Delaware, Newark, DE, USA
R O Maguire, North Carolina State University, Raleigh, NC, USA

Introduction

Animal manures are the solid, semisolid, and liquid by-products generated by animals grown to produce meat, milk, eggs, and other agricultural products for human use and consumption. They are mixtures of animal feces, urine, bedding materials (e.g., straw, sawdust, rice hulls), and other materials associated with animal production, such as waste feed, soil, wash waters, and any chemical or physical amendments used during manure handling and storage. Manures have been used as beneficial soil amendments since the dawn of civilization and were the primary soil amendment used in agriculture until the advent of chemical fertilizers in the 1940s. Today, manures continue to be regarded as valuable agricultural resources, because they are important sources of plant nutrients and are well known to improve soil physical and biological properties through the addition of organic matter. However, the changing nature of animal production in some countries, the move away from small farms that have adequate land bases to recycle manure nutrients through crop production, and the move toward geographically concentrated confined animal feeding operations (CAFOs) have raised concerns about the environmental and human health impacts of modern animal production and manure utilization practices.

Manure Generation, Use, and Properties

Summary of Global Trends in Animal Production

Animal production has increased worldwide in the past 40 years, although it is not distributed evenly over the globe (Figure 1a and Table 1). Long-term trends in animal production are most easily recognized if developed and developing countries are viewed separately, as greater intensification of animal production initially occurred in developed countries. Today, however, many developing countries are rapidly increasing animal production, while in most developed countries the number of animals produced is decreasing or leveling off, as seen for poultry production in Figure 1b. Calories from animal protein cost more to produce than calories from crops. Therefore as a general rule, as poorer countries develop and disposable income increases, the population can more easily afford animal protein, and there is a corresponding increase in animal production. As a result, meat consumption in developing countries is projected to double in the next 20 years, although it will still be below that of developed countries.

Globalization of agriculture has also changed the nature of animal production over the last few decades, increasing the distances that animal feed and animal products travel between sites of production and consumption. For example, poultry and swine produced in the eastern USA may use feed grains grown in the Midwestern US, with the meat

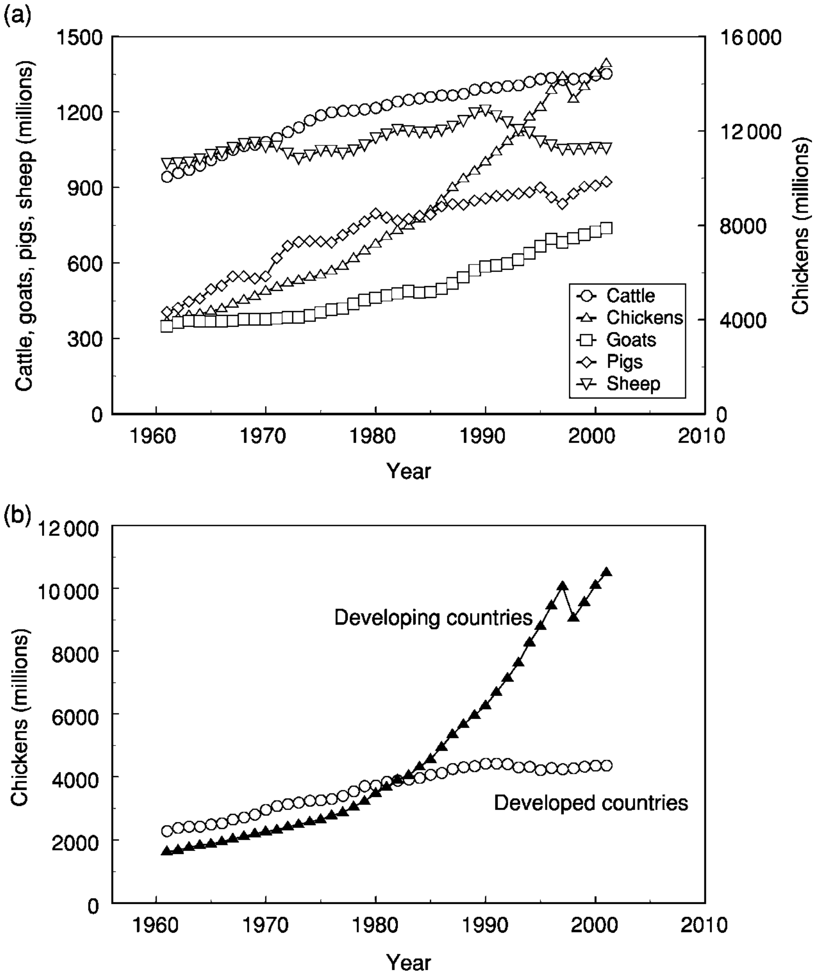


Figure 1 (a) Global trends in animal production, by major animal types (1961–2001); and (b) comparison of global poultry-production trends for developed and developing countries (1961–2002). Data from Food and Agriculture Organization of the United Nations (2003) Available online at: http://apps.fao.org/page/collections?subset-agriculture.

Table 1 Global animal production, by major animal species, for major geographic regions and for developed and developing countries, 2001

	Animal species				
Country	*Cattle (millions)*	*Chickens (millions)*	*Goats (millions)*	*Pigs (millions)*	*Sheep (millions)*
Asia	455	7180	462	551	371
Europe	101	1257	15	164	129
Africa	230	1276	219	18	250
North America	110	1988	1	72	8
South America	309	1765	22	55	75
Developed countries	326	4358	30	282	381
Developing countries	1026	10 501	708	641	675

Data from Food and Agriculture Organization of the United Nations (2003) Available online at: http://apps.fao.org/page/collections?subset-agriculture.

ultimately being consumed in Asia. In many countries these trends have also resulted in marked increases in the geographic intensity of animal production, with more animals being grown in confinement on farms that are located in close proximity to each other and that have the same, or declining amounts of cropland available for manure use. Figure 2 shows the change in the geographic distribution of confined animal production (all species) in the USA from 1982 to 1997.

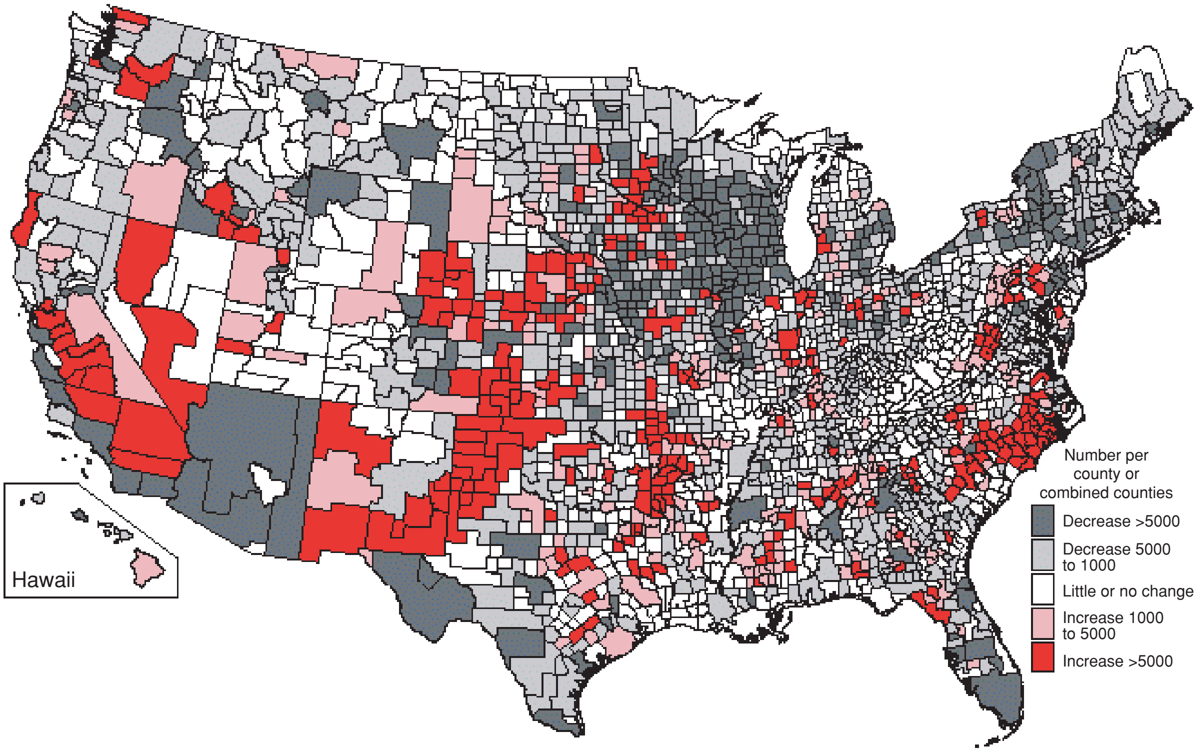


Figure 2 Change in the number of animal units in each county of the USA (1982–97). (Reproduced from Kellogg RL, Lander CH, Moffitt DC, and Gollehon N (eds) (2000) *Manure Nutrients Relative to the Capacity of Cropland and Pastureland to Assimilate Nutrients: Spatial and Temporal Trends for the United States*. Fort Worth, TX: GSA National Forms and Publication Center.)

Trends in Manure Production

The nature of the animal production system directly affects the amount of manure available for use as a soil amendment. For example, unlike the manure from animals grown in confinement (e.g., barns, feedlots), manure is not regularly collected from animals mainly grown on grassland or rangeland. Thus the amount of 'recoverable manure' that is available for land application as a soil amendment can vary widely between and within countries. However, as intensive cattle, swine, and poultry operations are projected to account for an increasing proportion of animal production in both developed and developing countries, collection and spreading of manure will probably become more and more common in the future. Manure production by species also differs widely by country; for example goats are very common in Asia and Africa, while cattle are more predominant in North and South America. In global terms, poultry production has grown the most in the last few decades, followed by pig and goat production, while the number of cattle produced has only increased slightly (Figure 1a). Indeed, in developed countries, poultry meat is seen as the only growth area toward 2010, in terms of per capita consumption. Despite these variations, the amount of manure produced by different animals is reasonably well known, as shown in Tables 2–4, which summarize manure production and characteristics for the major animal species raised today.

In general, the major trend today for animal, and hence manure, production in both developed and developing countries is for increasing intensification and consolidation of producers. This means that in developed countries, even though total manure production is leveling off, it is confined to fewer producers operating at fewer sites, with the result that the amount of manure generated in some regions is increasing markedly. For example, in the USA, the number of livestock operations decreased by 24% between 1982 and 1997, even though total animal production numbers were fairly stable. This concentration of animal production in specific areas in developed countries applies mainly to swine, cattle (beef, dairy, and veal), and poultry (chicken and turkey) production. In developing countries animal production is generally not as intensive but, as these countries develop, their production systems will inevitably become more intensive for the same economic reasons that promoted this trend in developed countries.

Table 2 Liquid-pit manure production and nutrient content for major animal species

Animal	Manure produced (kg year^{-1})	Nutrient concentration (kg 1000 l^{-1})			
		Total N	NH_3-N	P	K
Swine (farrow-finish)	17 007	3.4	1.9	1.3	2.3
Dairy (cow)	24 490	3.7	0.7	0.8	1.9
Dairy (heifer)	11 338	3.8	0.7	0.7	2.8
Beef (cow)	13 605	2.4	0.8	0.8	2.4
Poultry (layer)	59	6.8	4.4	2.7	3.3
Poultry (broiler)	38	7.6	1.6	2.1	2.9
Poultry (turkey)	128	6.4	1.9	2.1	2.9

Manure production values are expressed on a per-animal basis for beef, dairy, and swine and per bird space for poultry.

Adapted from Midwest Plan Service (2000) *Manure Characteristics. Midwest Plan Service*. Arnes, IA: Iowa State University Press.

Table 3 Lagoon-effluent manure production and nutrient content for major animal species

Animal	Manure produced (kg year^{-1})	Nutrient concentration (kg year^{-1})			
		Total N	NH_3-N	P	K
Swine (farrow-finish)	29 025	16	15	5	11
Dairy (cow)	41 270	21	19	4	12
Fattening cattle	19 955	12	11	4	10

Manure production values are expressed on a per-animal basis for beef, dairy, and swine and per bird space for poultry.

Adapted from Midwest Plan Service (2000) *Manure Characteristics. Midwest Plan Service*. Arnes, IA: Iowa State University Press.

Table 4 Solid-manure production and nutrient content for major animal species

Animal	Manure produced (kg year^{-1})	Nutrient concentration (kg Mt^{-1})			
		Total N	NH_3-N	P	K
Swine (farrow-finish)	7773	5.8	2.5	1.4	1.7
Dairy (cow)	12 698	4.1	0.8	0.5	2.0
Dairy (heifer)	5896	4.1	0.8	0.5	2.4
Beef (cow)	6077	2.9	1.2	0.7	2.4
Poultry (layer)	18	14.0	4.9	9.2	8.9
Poultry (broiler)	8	18.9	4.9	9.5	12.3
Poultry (turkey)	21	16.5	3.3	9.0	10.2

Manure production values are expressed on a per-animal basis for beef, dairy, and swine and per bird space for poultry.

Adapted from Midwest Plan Service (2000) *Manure Characteristics*. Arnes, IA: Iowa State University Press.

In some cases, the geographic intensification of animal production has led to manure generated within confined areas of production, containing nutrients in excess of local crop requirements, while other areas within the region or country have little or no local supply of manure nutrients for crop production. This is because the economic value of the manure cannot support the costs of its redistribution to areas where it may be needed as a soil amendment. As illustrated for the USA in Figure 3, this has resulted in large 'surpluses' of manure nutrients in some areas where animals are produced and increasing concerns about nonpoint source pollution of air, soil, and water in areas where animal production is concentrated. These concerns arise because farmers who have excess manures, relative to what is needed to produce crops, generally have no economically viable alternatives to land application. Because of this, manure nutrients are often overapplied or applied at inappropriate times of the year, leading to increased nutrient losses to surface water and groundwater by runoff and leaching, and to the atmosphere by volatilization. In some developed countries (e.g., the Netherlands), regulations have been enacted to decrease animal production intensity because of the environmental impacts associated with the disposal of excess manures. Efforts to control manure generation and use are expected to increase in developed countries in the future; however, in developing countries this may not occur until manure and nutrient excesses begin to create similar environmental problems.

Manure Characteristics by Species

The physical properties and nutrient content of animal manures are reasonably well known and tables summarizing manure characteristics have been developed for use by farmers and their advisors (Tables 2–4). Values such as these, however, are only guidelines, as it is recognized that manures can vary widely in composition both between and within species. Therefore, regular testing of manures for properties that influence their agricultural value and potential environmental impacts is often advised. Results from manure testing have consistently shown that differences in manure nutrient content between animal species are apparent (Tables 2–4). For example, both N and P concentrations in manure increase in the order cattle < swine < sheep and goats < poultry. Poultry litter contains two to four times the concentration of manure nutrients as other livestock manures, which leads to particular problems in areas where the poultry industry has intensified, compared with areas of other intensified livestock industries. For example, in the USA, of the 73 counties that

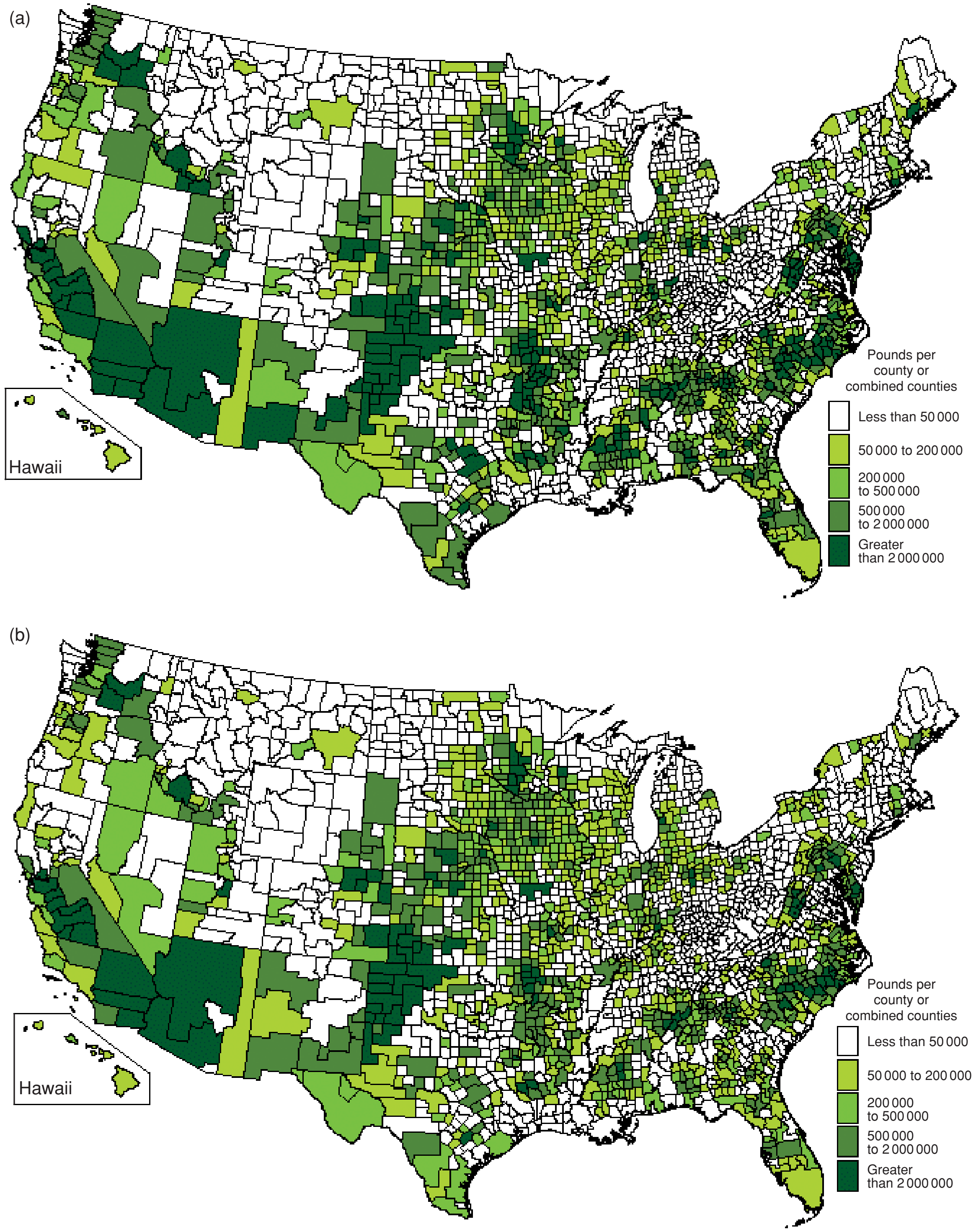


Figure 3 Potential excess of manure (a) nitrogen and (b) phosphorus for each county of the United States in 1997, assuming no export of manure from farms. (Reproduced from Kellogg RL, Lander CH, Moffitt DC, and Gollehon N (eds) (2000) *Manure Nutrients Relative to the Capacity of Cropland and Pastureland to Assimilate Nutrients: Spatial and Temporal Trends for the United States*. Fort Worth, TX: GSA National Forms and Publication Center.)

had insufficient land for the amount of manure N generated, 82% had poultry as the dominant livestock, while of the 160 counties with excess manure P, 64% were dominated by poultry production (Figure 3).

The variability in manure composition is due to a number of factors, including animal species and age, diet, digestibility of feed materials, environmental conditions in animal housing facilities, and the means by which manures are handled and stored prior to use or disposal. For example, the amount of feed concentrates and other feed additives (e.g., enzymes) in a diet (and thus nutrients in manure) can vary between developed and developing countries; manure storage and drying can reduce nitrogen (N) content through ammonia volatilization; and manure collected from open-housing areas or stored in lagoons can be diluted by rainfall.

Major Uses of Animal Manures

Land application to meet crop N, P, and other nutrient requirements is by far the most widespread and longstanding use of animal manures. However, other uses do exist and are becoming more common, especially in areas where animal production has become so intensive that the amount of nutrients in the manure produced exceeds the nutritional requirements of crops grown on the local land base. Sometimes manure is fed back to animals as a cheap nutrient source (e.g., feeding chicken litter to beef cattle). However human and animal health concerns have been raised about this practice, in the wake of the 'mad cow disease' outbreak in Europe (caused by feeding contaminated animal products back to cattle), and this is not viewed today as a major alternative to land application of manures. Mushroom production in certain areas can use large amounts of manure, especially from horses and poultry. In China, manures are used extensively in fish production. Several other options exist for manure use, and although these uses are only carried out on a relatively small scale today, they represent a growth area, as increasing intensification of agriculture produces excess amounts of manure in localized areas of more countries. Examples of other manure uses are composting to produce high-value soil amendments, methane generation via anaerobic digestion, burning to generate energy, and pelletizing manures to make organic fertilizers.

Best Management Practices (BMPs) for Animal Manures as Soil Amendments

Overview

Years of research have identified the most effective approaches to manage animal manures as agricultural soil amendments and many references provide detailed discussions on this subject. In general, manures have historically been viewed as organic fertilizers and have been applied at rates that would provide adequate nutrients for crops to be grown on the farm, based on realistic yield goals and the results of soil and manure tests that quantify the amount of manure nutrients needed to attain economically optimum yields. In conjunction with this, soil and water conservation practices (e.g., buffer strips, grassed waterways, and cover crops) are usually recommended to prevent off-site movement of manures or manure nutrients. Other benefits of manure use (e.g., increasing soil organic matter content and improving soil physical and biological properties) are recognized but usually do not serve as the basis for determining manure application rates. The most effective BMPs for manure use vary with the type of agriculture being practiced. This can be illustrated by comparing two of the most predominant agricultural settings worldwide: grassland agriculture and arable crop production.

Manure management for grassland agriculture In general, there are two basic systems associated with grassland agriculture: (1) pasturing, and (2) confinement. Animals such as cattle or sheep graze in pastures and deposit manure as they produce it. Thus, manure 'application' is somewhat random and occurs without any form of incorporation into the soil. Where animals are confined indoors for at least part of the year, the manure they generate is collected before being land-applied. For confinement systems, manures are most commonly applied to the surface of pastures without tillage, although in some cases they may be injected into the soil using specialized equipment. With intensification of grassland agriculture, there has been a trend toward animal confinement and manure spreading, and away from solely pasturing animals. It is now estimated that about half of the manure produced in the USA can be collected, with much of the remainder being directly deposited by grazing animals. As more manure is collected, the opportunities to implement BMPs that better utilize the collected manures as soil amendments, or for other end uses, increase.

In general, manures are valuable sources of nutrients for grass and animal production in pasture systems. However, because they are surface-applied, several pathways for the loss of nutrients and other manure constituents (e.g., pathogens, hormones, and antibiotics) exist that are more pronounced than when manures are incorporated into arable soils. These nutrient losses are both economic (loss of plant nutrients) and environmental (deleterious impacts on air and water-quality) problems. For example, substantial gaseous losses of N by volatilization of ammonia occur directly from manures lying on the soil surface. This may increase the need to purchase N fertilizers to compensate for the lost manure N and also contribute to environmental concerns when the volatilized NH_3 is redeposited on to soils (which may cause soil acidification) or into aquatic ecosystems (where the N may contribute to eutrophication).

As grassland is unlikely to be tilled regularly, there is also a tendency for nutrients to accumulate very near the soil surface as manures slowly decompose. This is particularly true for P, which is strongly adsorbed by soil clays and oxides. When runoff occurs, dissolved and colloidal P either in the manures lying on the surface or contained in the nutrient-rich uppermost layers of the grassland soil can be transported to surface waters, increasing the potential for eutrophication of surface waters.

Managing manures in grassland agriculture thus focuses on the use of BMPs that sustain optimum grass production and also prevent surface and gaseous losses of nutrients and other manure constituents. Examples include the use of soil testing and equipment calibration to prevent overapplication of manures; avoiding poorly timed and badly placed manure applications (e.g., immediately before rainfall events or snowmelts and in close proximity to surface waters or wellheads); use of chemical and physical manure amendments that decrease the solubility or volatility of N and P; and injection of manures below the grassland soil surface, although this is a practice that entails extra technology and costs and is not widely used at present.

Managing manures for arable crop production Unlike manures in grassland agriculture, animal manures used in arable crop production are commonly incorporated into soils by some form of tillage (e.g., plowing, disking, rototilling) or by injection. This is not always the case, however, as many agricultural crops today are produced using reduced or ‘no-till’ systems, where the soil is not plowed annually or is only tilled to a shallow depth to prepare a seedbed for annual crops. Manures incorporated into soils in arable cropping systems decompose as a function of soil microbial activity, which in turn is dependent upon soil temperature, moisture, aeration, and soil chemical properties (e.g., pH, salinity, nutrient status). As the manures decompose, they release nutrients into the soil solution that can be retained by chemical processes (e.g., ion exchange, adsorption, precipitation) or as part of the soil biota (e.g., incorporation into microbial biomass). Manure decomposition can also affect soil pH, soluble-salts concentration, and soil organic matter content, all of which affect plant growth and the potential for nutrient losses. Much research has been directed toward the development of predictive models that can estimate the rate of manure decomposition in soils and the total amount of nutrients released from different types of manures as a function of soil properties and environmental conditions.

As with grassland agriculture, manures incorporated into arable soils are well known to build soil fertility and improve soil physical properties such as soil structure and tilth and thus are important soil amendments for crop production. Nutrient losses can also occur from arable soils, with the greatest concerns being the erosion of soil particles (and nutrients that are bound to these particles, particularly P) and the leaching of soluble nutrients, especially nitrate-N, through the soil profile to groundwater. Soil erosion concerns are greater with arable crops than grasslands because of the tillage practices associated with manure incorporation and the lack of the permanent plant cover that can filter eroding soil particles from runoff. Gaseous losses of N are much reduced when manures are incorporated into arable soils, because manure ammonium-N is bound to soil colloids and is also not in direct contact with the atmosphere. Similarly, the potential for losses of soluble P is less in arable soils, because manures are not lying on the soil surface and because tillage prevents the development of the highly enriched zones of P accumulation often found in the upper few centimeters of untilled grassland soils.

Manure management for arable crops is similar to that described above for grasslands, with a few key differences. There is a much greater focus on controlling soil erosion and there are greater limitations on the times of year that manure can be applied to annual row crops. In many cases manures must be applied well before planting the crop, which then increases the likelihood for nutrient losses.

Contemporary Approaches to Manure Management

As animal agricultural operations have become increasingly intensive, many problems have emerged that point to the need for improvements in manure-management practices. While these problems are more serious in developed countries, where animal production is most intensified, they are emerging issues in developing countries, where information on how to avoid the environmental problems associated with intensive animal agriculture will be needed in the future.

One of the main problems with traditional approaches to manure management is the fact that, while manures provide most of the nutrients needed by crops, the relative proportions of plant nutrients present in manures are not in balance with the amounts needed by many crops. This can result in under- or overapplication of some nutrients. For example, manures have usually been viewed as N sources and applied to meet crop N requirements. However, because the N:P ratio in manures is not in balance with the N:P ratio in most crops, applying manures to meet crop N needs, long the recommended practice, has often resulted in the overapplication of P and the buildup of P in soils to values well above those needed for optimum crop yields. Because of this, concerns now exist in many countries about losses of P to surface waters and shallow groundwaters from soils that have received long-term manure applications. Other emerging concerns for manure use as soil amendments include environmental impacts of other manure constituents such as pathogens, hormones, antibiotics, and nonessential trace metals on human health and environmental quality.

In some countries, these concerns have resulted in the development of more-intensive, structured, and even regulatory approaches for manure management, such as the comprehensive nutrient management plans (CNMPs) now recommended by the US Natural Resources Conservation Service for animal feeding operations (AFOs). The six basic components of a CNMP are described below and may apply, in principle, not only to the intensive animal agriculture common in the USA, but to animal production operations worldwide, including the less-intensive systems in many developing countries.

1. Manure and wastewater-handling and storage: This element of the CNMP focuses on the activities at the animal production facility (e.g, barn, feedlot) and any areas where manure is handled and stored prior to land application. The general goals are to minimize the amount of waste produced and to handle, store, treat, and transfer wastes to off-site uses in a manner that minimizes any losses, spills, or direct discharges to the environment. Areas of particular concern include providing adequate manure storage, preventing direct losses of pollutants from the farmstead and manure storage areas to nearby surface waters, minimizing odors and air-quality concerns, and safely disposing of dead animals, spoiled feed, and the like;
2. Land treatment practices: Because land application is the main use of most animal manures, well-designed land-treatment practices are vital components of a CNMP. Factors that must be addressed are: selection of the appropriate rate, method, and timing of application of manures, relative to crop needs; implementation of soil conservation practices that prevent soil erosion and the loss of soluble nutrients from manures and manure-amended soils; and establishment of monitoring programs to assess the effectiveness of the land-treatment practices over time;
3. Nutrient management: This refers to the development of a systematic approach to manage all nutrient sources on the farm, not just manures. Typically this begins with an effort to maintain nutrient balance in the operation, where nutrient inputs in feed, fertilizers, animals, and other sources do not exceed nutrient outputs in crops and animal products, except as needed to build soil fertility to optimum levels for crop growth. Key factors considered in this component of the CNMP include a farm-scale nutrient budget (i.e., is there a nutrient surplus or deficit on the farm?), crop nutrient requirements at realistic yield goals, soil-testing to assess the need for additional nutrients, manure-testing to determine the quantity of manure nutrients available and the need for off-farm nutrients, and the most efficient means to apply all nutrient sources to crops on the farm. Other considerations include how the approach used to manage nutrients might affect air quality, odors, pathogens, and the buildup of soluble salts and heavy metals in soils;
4. Record-keeping: Effective implementation of a CNMP requires good record-keeping. The following information should be recorded, by field, and reviewed to assess current practices and detect trends with time: basic farm characteristics (e.g., amount of land; weather conditions; number and type of animals or animal products produced; type and acreage of crops grown; amount of manure produced, used on the farm or transferred off-site); manure storage, handling, and treatment; soil, manure, and plant tests; crop-management practices (hybrid, planting date, seeding rate, irrigation, yields), and nutrient application practices (source, amount applied, when applied, method applied, equipment calibration);
5. Feed management: Animal-based agricultural operations are increasingly investigating the opportunity to improve nutrient use efficiency by animals, which, in turn, reduces the excretion of nutrients such as N and P. Research has shown that modifying animal diets by reducing the concentration of mineral supplements, or by including enzymes and other feed additives, can result in the same level of animal performance but reduce the amount of N and P excreted. This results in both an economic savings (fewer nutrients are needed for the same level of production) and lower nutrient concentrations in manures, which can be of benefit in areas where nonpoint nutrient pollution of surface and ground waters is of concern;
6. Alternative uses for manures: There is growing interest in developing alternatives to land application for some types of animal manures, particularly in areas where animal production has been geographically intensified and nutrient surpluses exist (see Figure 3). To sustain, or expand, operations in areas

such as these, it is necessary to find alternatives to land application, and this is a central part of an efficient CNMP. This can generate additional income and also help farmers redistribute surplus nutrients to off-site uses. The most common approaches investigated to date have been energy production (methane generation and incineration for electricity), production of value-added products such as composts and pelletized fertilizers, and transportation of manures from areas of nutrient surplus to nutrient deficit, usually mediated by ‘manure brokers.’

Summary

The use of animal manures as soil amendments must be done in a comprehensive manner to maximize their agricultural value and minimize their potential impacts on environmental quality and human health. There is no doubt that animal manures are valuable natural resources, particularly in countries where the availability of inorganic fertilizers is limited. When managed properly, manures provide plant nutrients, build soil organic matter, and improve soil physical properties, all of which are vitally important for soil quality and crop production. At the same time, the growing global trend toward intensification of animal production is gradually concentrating manure generation on smaller amounts of land. Transportation costs and the lack of an organized, efficient manure processing and distribution infrastructure are the major limitations to relocation of manures from areas where they are in surplus to areas where they are badly needed. When an adequate land base is not available to recycle manure nutrients in a balanced manner, overapplication of nutrients such as N and P, which can negatively affect air, soil, and water quality, becomes a potential problem. Regional- and/or national-scale planning is required to develop and implement programs that can sustain modern animal production and use animal manures in an environmentally sound manner.

Further Reading

Beegle DB, Carton OT, and Bailey JS (2000) Nutrient management planning: justification, theory, practice. *Journal of Environmental Quality* 29: 72–79.

Food and Agriculture Organization (1995) *World Agriculture: Towards 2010*. New York: John Wiley.

Food and Agriculture Organization (FAO) (2002) http://www.fao.org.

Hatfield JL and Stewart BA (eds) (1998) *Animal Waste Utilization: Effective Use of Manure as a Soil Resource*. Chelsea, MI: Ann Arbor Press.

Kellogg RL, Lander CH, Moffitt DC, and Gollehon N (eds) (2000) *Manure Nutrients Relative to the Capacity of Cropland and Pastureland to Assimilate Nutrients: Spatial and Temporal Trends for the United States*. Fort Worth, TX: GSA National Forms and Publication Center.

Midwest Plan Service (2000) *Manure Characteristics. Midwest Plan Service*. Ames, IA: Iowa State University Press.

Sims JT (1994) Animal waste management. In: *Encyclopedia of Agricultural Sciences*, pp. 185–201. New York: Academic Press.

Sims JT (1995) Organic wastes as alternative nitrogen sources. In: Bacon PE (ed.) *Nitrogen Fertilization in the Environment*, pp. 487–535. New York: Marcel Dekker.

Steele K (1995) *Animal Wastes and the Land–Water Interface*. New York: Lewis Publishers/CRC Press.

MARBUT, CURTIS FLETCHER

J P Tandarich, Hey and Associates, Inc., Chicago, IL, USA

The brilliance that characterized the life of Curtis Fletcher Marbut, dedicated to soil science in the service of humanity, ended on August 25, 1935 in what was then Harbin, Manchuria, China. He was the best-known soil scientist–geographer of his day, and his reputation was worldwide. Marbut was a pioneering soil scientist in the classroom, office, field, and laboratory. More than that, he was a disseminator of ideas; an internationalist who did not let politics, ideology or his personal health and welfare stand in the way of scientific advancement and friendship. The biography of Marbut is a remarkable story of a ‘self-made’ man who came from a progressive Ozarks family in rural southwest Missouri to become one of the foremost agricultural geologists, geographers, and soil scientists in the world during the formative period of modern soil science in the early twentieth century. Modern soil science has its origins in late eighteenth- and early nineteenth-century agricultural chemistry and agricultural geology. The term Bodenkunde, coined in 1837 by German agricultural chemist Carl

Sprengel, meant soil knowledge – a collection of facts. Friedrich Albert Fallou, a German agricultural geologist, attempted to elevate Bodenkunde to a formal science of soil that he called Pedologie. The German scientists never adopted the term, preferring to use Bodenkunde to mean soil science. Russian agricultural geologist Vasilii V. Dokuchaev used the word 'pochvovedenie' for soil science. In the twentieth century, pedology came to mean the study of the genesis, morphology (appearance), nature, classification, and geographic distribution of soil.

This short narrative summarizes the highlights of Marbut's life as a native of the Missouri Ozarks, an academic, a scientist with the US Soil Survey, an explorer and internationalist, and a scientist with a Unitarian philosophy of service to humanity for which he ultimately gave his life.

Missouri Ozarks Native and Academic

Marbut was born during the American Civil War, on July 19, 1963, to Nathan and Jane Browning Marbut, along Little Flat Creek near Verona in Lawrence County, Missouri. After surviving whooping cough in the first weeks of life, he grew up in an atmosphere of love, learning, and a concern for others, largely fostered by his Quaker mother.

His school attendance began in the one-room Marbut family school, and continued at a school in McDowell, Missouri, where he also taught when he was 17. Using his teacher's earnings, he enrolled in the Collegiate Institute of Professor Noah Lee Maiden at Cassville, Missouri (Figure 1), an experience that led him ultimately to enroll in the University of Missouri at Columbia (UMC) in 1885. He went on to receive a BS degree from UMC in 1889 and, in 1890, was appointed to the Missouri Geological Survey. In 1891, he married Florence L. Martin of Cassville, Missouri (Figure 2) and established residence in Jefferson City, Missouri. They had five children: Louise Marbut Moomaw (Figure 3), Thomas Fiske Marbut, William Martin Marbut, Helen Marbut, and Fredrick Browning Marbut. The lifelong correspondence between Louise and her father gives an insight into his life not possible through his published works alone.

The family moved to Cambridge, Massachusetts, in 1893 so that Marbut could enroll as a graduate student in geology at Harvard University. Although his major professor and adviser was the eminent pioneer geographer and geomorphologist William Morris Davis, Marbut revealed in his correspondence to Louise that Nathaniel Southgate Shaler, a pioneer geographer and agricultural geologist, was his academic and professional inspiration to study soils.

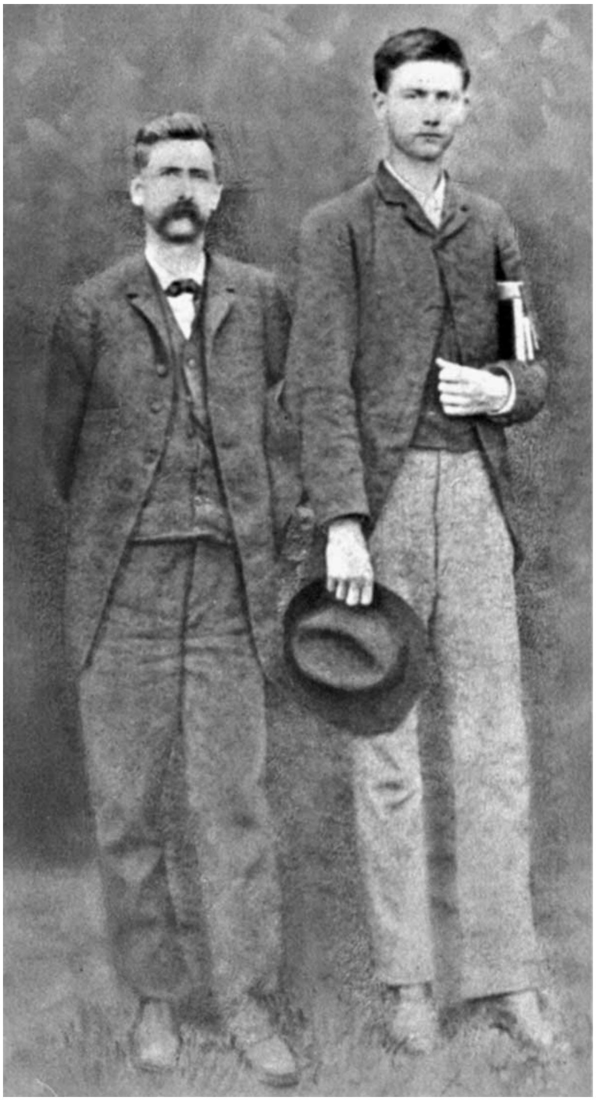


Figure 1 Professor Noah Lee Maiden and Curtis Marbut, 1880. Curtis Fletcher Marbut Papers, 1852–1963, Western Historical Manuscript Collection, Columbia.

Just prior to and during Marbut's tenure at Harvard, Shaler had become highly regarded for his publications on soils and natural resources conservation. The most notable of these were *Animal Agency in Soil-Making* (1888), *The Origin and Nature of Soils* (1891), *Nature and Man in America* (1891), *The Interpretation of Nature* (1893), and *The Economic Aspects of Soil Erosion* (1896). Shaler was a compelling lecturer who promoted the education of the whole student: mentally, physically, and spiritually. Marbut received an MA in 1894 under Davis. In 1895, he completed the residence requirements for the PhD and returned with his family to the UMC with a faculty appointment. Although he finished and published his PhD dissertation in 1896 with Davis'

Figure 2 Florence Martin Marbut, *c.* 1890. Western Historical Manuscript Collection, Columbia.

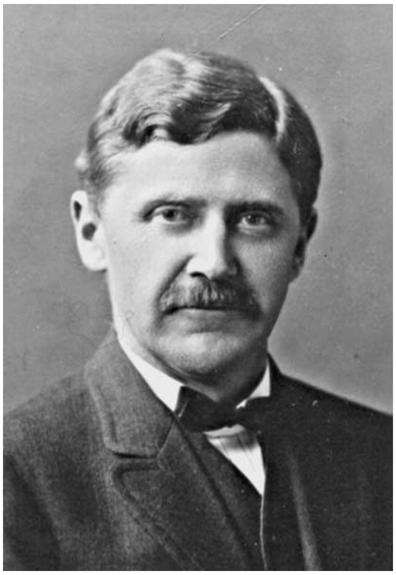


Figure 4 Merritt Finley 'M.F.' Miller, *c.* 1910. Personal collection of John P. Tandarich.

Figure 3 Louise Marbut Moomaw, *c.* 1920. Western Historical Manuscript Collection, Columbia.

approval, he never returned to Harvard for his oral defense exam and the PhD was never awarded.

Marbut served the UMC Geology Department as instructor (1895–1897), assistant professor (1897–1900), and professor (1900–1910), and was simultaneously appointed assistant geologist in the Missouri Geological Survey until 1904. His PhD dissertation *Physical Features of Missouri* published in 1896 by the Missouri Geological Survey was a pioneering study in physiography and surface hydrology. While at the UMC, Marbut published numerous studies, including: erosional remnants Crowley's Ridge, Cote Sans Dessein, and Grand Tower (1898); coal stratigraphy of areas of northern Missouri (1898); the lowlands of Southeastern Missouri (1902); the geology of Morgan County, Missouri (1908); and the soils of the Ozark region (1910).

In 1904, Marbut and Agriculture College Dean Henry J. Waters collaborated on a large-scale plaster physiographic map of Missouri that won a gold medal at the 1904 Louisiana Purchase Exposition in St Louis. In addition, Marbut and Waters coauthored the first attempt at a general soil map of Missouri that, together with two geological maps and a chapter on geology and physiography by Marbut, were published in the Exposition's volume, *The State of Missouri*. The following year, due to the influence of Waters and Cyril G. Hopkins of the University of Illinois, the University of Missouri Board of Curators organized the Missouri state soil survey and named Marbut as director. For the next 5 years, Marbut oversaw and contributed to the preparation and publication of individual county soil surveys. Thus began a lifelong interdisciplinary collaboration between Marbut and UMC faculty members Merritt Finley (known as M.F.) Miller of the Agronomy Department (Figure 4) and Homer L. Schantz of the Botany Department (Figure 5).

When, in 1909, USDA Bureau of Soils Chief Milton Whitney came to the University as keynote speaker at the annual Farmer's Week activities, he lured Marbut into the federal service as Special Agent in the national soil survey. Marbut's first task was to make a reconnaissance soil survey of the Ozark region of

Figure 5 Homer L. Schantz and Curtis Marbut taking soil samples in South Dakota, 1925. Western Historical Manuscript Collection, Columbia.

Arkansas and Missouri. However, Marbut's wife Florence died from pneumonia that same year, leaving him in care of the family. When he was promoted to Scientist in the Soil Survey in 1910, he took a 2-year leave of absence from UMC for him and the family to take up residence in Washington, DC.

Scientist in Soil Survey

Marbut began his federal professional career by publishing the results of his Ozark soil survey in 1911, and leading the preparation of a monumental tome on US soils published in 1913. Marbut was dissatisfied with the latter work because the soil concepts and descriptions used in the soil survey program were unsystematically applied throughout the country. Nevertheless, he was appointed Scientist-in-Charge of the Soil Survey in 1913, but he and the family returned to the UMC. Suddenly, he decided to return to Washington and left the family at home in Columbia with a housekeeper and Louise to begin the most significant part of his professional life.

Upon his return to Washington, Marbut revisited the soil survey concept–description problem. He wrote to Louise about the situation that, "For four years I continued to grope in darkness without a ray of light." Then in 1914, Marbut discovered one of the two copies of *Die Typen der Bodenbildung* by Konstantin D. Glinka, a student of pioneer Russian soil scientist Vasilii V. Dokuchaev. The books were obtained by the USDA Departmental Library just prior to the outbreak of World War I. After his first reading of Glinka, Marbut wrote, "it comes to me suddenly – the soil profile." He translated the work in 3 years, and then unsuccessfully tried to publish it 1920 in during the height of the 'Red Scare' of anti-Bolshevik sentiment. Another obstacle was Whitney's disdain for foreign ideas and work. As with the departure of his predecessor Charles Nelson Coffey, Marbut feared a similar fate awaited him from Whitney who "had a reputation for getting rid of those who did not agree with him." When confronted by Whitney, Marbut told him that he would "follow his policies, but would think for myself." Marbut encountered anti-Bolshevik sentiment in the Bureau of Soils, writing: "I was called a Russian worshipper by my own men."

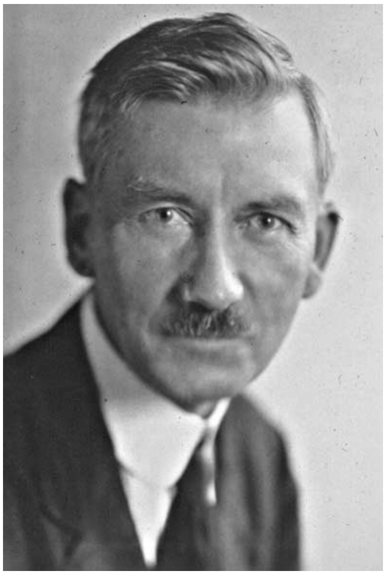


Figure 6 Curtis Marbut, *c.* 1920. Western Historical Manuscript Collection, Columbia.

Marbut evaded this ethnocentric attitude, as he said, by having "to work in secret." By 1920, he had prepared a limited number of carbon copies of the Glinka translation for circulation to a few selected colleagues (Figure 6). In 1921, at the American Soil Survey Workers Association (ASSA) meeting in East Lansing, Michigan, he introduced the Russian concepts of classification and soil profile without identifying their origin. The ASSA, founded in 1920, was an organization started by scientists in soil survey for the promotion and publication of their ideas. Roy W. Simonson, however, has stated that the ASSA was formed "to counteract the growing influence of C.F. Marbut." Nevertheless, the ASSA was influential in disseminating ideas and it joined with the Soils Section of the American Society of Agronomy to create the Soil Science Society of America (SSSA) in 1936.

He worked with ASSA colleagues and soil survey parties to apply these 'new ideas' in the USA. Marbut's own notes in his personal papers clearly indicate his dedication to convince US soil scientists of the efficacy of using the Russian profile concept as a basis for soil descriptions. To this end, he personally visited as many soil survey field parties as possible and participated in their yearly field reviews to exercise his influence and powers of persuasion – a practice that he continued his entire career.

During this time, Marbut's reputation grew in disciplines outside soil science, such as Pleistocene geology. An example of this is his interaction with University of Illinois faculty member and Illinois state geologist Morris M. Leighton. In the 1920s, Leighton was working to develop the weathering profile concept, originally conceived by Iowa state geologist George F. Kay. He stated that both Glinka's book and Marbut influenced him during the latter's visits to the University of Illinois between January and July 1923. Leighton credited Marbut for further insights obtained on the soil profile and its homologous application in geology as the weathering profile.

By the 1924 ASSA meeting, Jethro Otto Veatch of Michigan State University knew of the Russian origin of the ABC soil profile when he publicly referred to it as 'the Glinka scheme.' The foreign source of this soil concept was known among the soil scientists and seemed to be less of a problem.

Marbut's 1927 presentations at the First International Congress of Soil Science (see below; Figure 7) and his 1928 lectures in the Graduate School of the USDA represented further opportunities to promote the Russian ideas. The lectures, some of which were published in 1951, contained numerous references to the Russian work and influenced many soil scientists, particularly his own staff members, such as Mark Baldwin, Charles Kellogg, and James Thorp (Figure 8). The results of his lifelong work on the description and classification of the soils of the USA were published in the 1935 *Atlas of Agriculture* that included a monumental soil map.

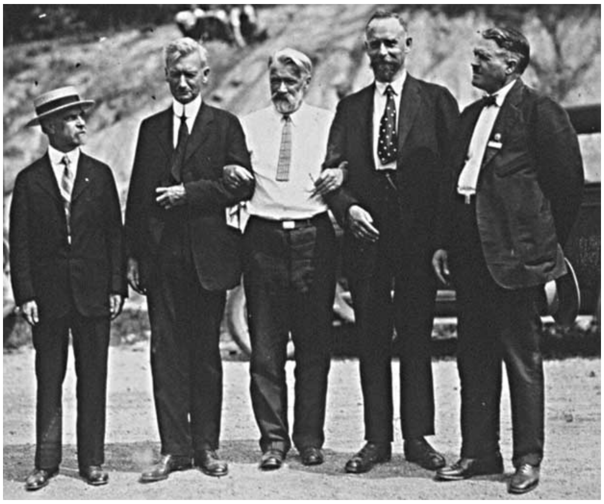


Figure 7 Oswald Schreiner, Curtis Marbut, Konstantin D. Glinka, D.J. Hissink, and Arthur G. McCall, 1927. Western Historical Manuscript Collection, Columbia.

Internationalist–Explorer

From 1920 until 1935, Marbut assumed the role of an internationalist in the field of soils. His extensive travels and explorations during this period carried him to many parts of the world and gained for him the well-earned reputation as an international authority on soil geography – as much at home in the study of the soils of Europe, Africa, and South America as those of his homeland (Figures 9 and 10).

Marbut acquired his first-hand knowledge of soils and landscapes of the world through his exhausting schedule of trips and expeditions. His first trip to Europe was on a sabbatical from UMC in 1899–1900 with colleague Isidor Loeb, during which they studied the Alps, the Hungarian plain, the Mediterranean region, the British Isles, and Scandinavia. In

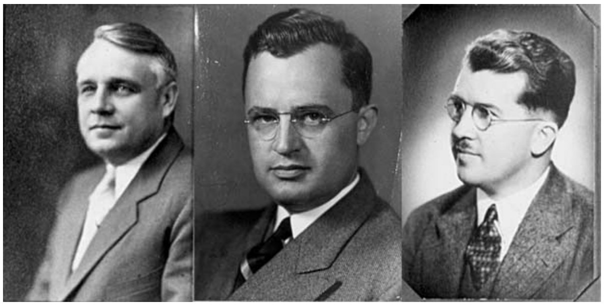


Figure 8 Mark Baldwin, Charles E. Kellogg, and James Thorp, 1937. Personal collection of John P. Tandarich. Baldwin and Thorp Portraits originally courtesy of Mrs Mark (Mary) Baldwin and Cyril Harvey family.

Figure 9 During the Amazon expedition, 1923. Western Historical Manuscript Collection, Columbia.

1918, Marbut and Schantz, who had gone from the UMC to the USDA Bureau of Plant Industry, studied the vegetation and soils of Africa for the American Commission to Negotiate Peace. In April 1919, Marbut and Bureau of Soils colleague Hugh Hammond Bennett went to Latin America to soil survey a disputed boundary between Guatemala and Honduras, sponsored by the American Geographical Society (AGS). He traveled to Prague, Czechoslovakia, in 1922 for the International Agrogeology Conference and subsequent field trip. From July 1923 to January 1924, he mapped the interior Amazon basin as a representative of the Brazilian-American Rubber Commission. Immediately afterward, until July, he studied the soils of Argentina under sponsorship of the AGS. Later that year, he traveled to Rome for the meeting of agro-geologists at which he assisted in the formation of the International Society of Soil Science (ISSS). At this meeting, Jacob G. Lipman of Rutgers University and he began preparations for holding the First International Congress of Soil Science in Washington, DC in 1927.

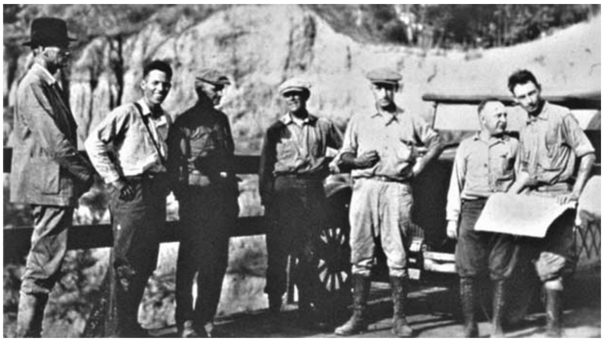


Figure 10 Curtis Marbut (far left) visiting soil survey party in North Dakota in early 1920s. Western Historical Manuscript Collection, Columbia.

Though Marbut made two presentations on soil classification during the Commission V session of the 10-day First International Congress that lasted from 13–22 June, 1927, his legacy of the meeting was the organization of the post-meeting Transcontinental Excursion (Figures 11 and 12). The Excursion was a 12 000-mile railroad train trip of 30 days (commencing on June 22) through 23 states in the

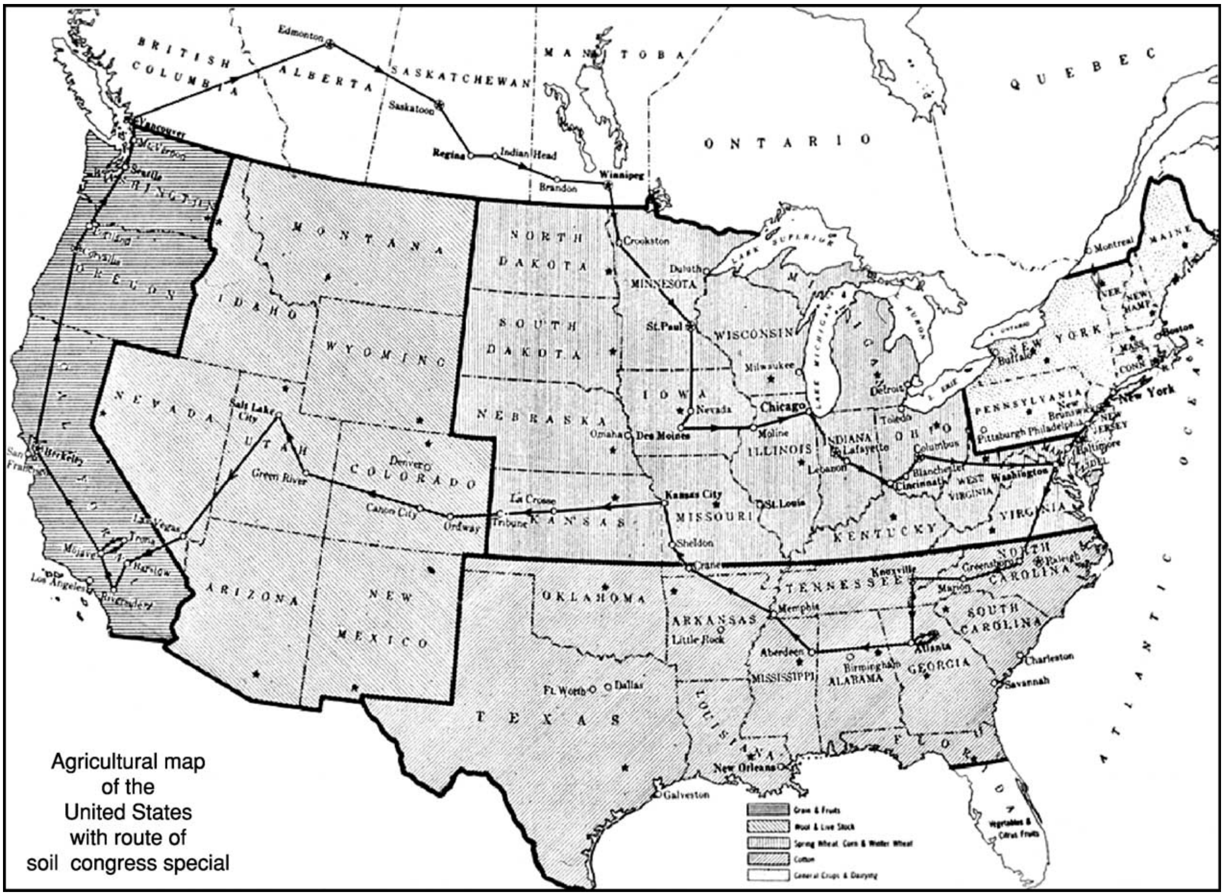


Figure 11 Route of the Transcontinental Excursion train ‘The Soil Congress Special’ through the USA and Canada, 1927. Archives of John P. Tandarich.

Figure 12 Typical Transcontinental Excursion stop to examine a soil pit, 1927. Marbut is standing third from right. Western Historical Manuscript Collection, Columbia.

Figure 13 Curtis Marbut (second from left) with soil scientists in Kansas, 1932. Western Historical Manuscript Collection, Columbia.

USA and four Canadian provinces to examine soils and landscapes at 34 stops. The Excursion advanced the sharing of ideas among soil scientists from around the world through the constant opportunities for interaction on the train, at the individual stops and through the daily Excursion newspaper, the *Boden Bull.*

In early 1928, Marbut visited Bennett's soil survey party in Cuba. In May 1929, he went to Danzig (modern Gdansk) for ISSS Commission V meetings and then to Soviet Russia in mid-1930 for the Second International Congress of Soil Science. The Excursion following the Second Congress meetings gave Marbut a view of the soils of western Russia from Moscow south to the region between the Black and Caspian Seas. He also visited Germany, particularly the Black Forest region, immediately afterward. James Thorp reported that Marbut visited his soil survey party in Puerto Rico early in 1932. This was immediately after attending the Fourth International Congress of Sugar Cane Technologists (Figures 13 and 14).

In 1933, Marbut accepted an assignment to advise the National Geological Survey of China in their organizing of a soil survey. To this end, Marbut sent James Thorp, his colleague from the Washington, DC soil survey office, to China in 1933 to train students in soil survey methods. By early 1935, enough progress in the soil survey of China had been made that it was necessary for Marbut to travel to China to render opinions on the descriptions and classification of the soils. In consultation with Thorp, Marbut decided to travel to China following his attendance at the Third International Congress of Soil Science in Oxford, UK (Figure 15).

In August 1935, Marbut traveled to England as a delegate to the Third Congress. Immediately after the

Figure 14 Participants in Second International Congress of Soil Science at Leningrad, 1930. Marbut is second from left in front. Western Historical Manuscript Collection, Columbia.

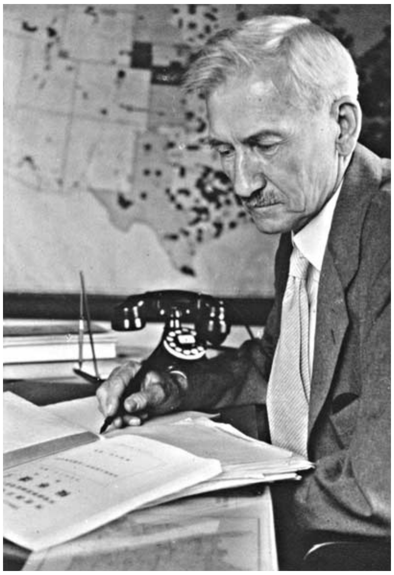


Figure 15 Curtis Marbut preparing for trip to China, 1935. Western Historical Manuscript Collection, Columbia.

meetings, he traveled through Europe and then, by the Trans-Siberian Railroad, through southern Russia. During the Siberian leg of the trip, Marbut was stricken with double pneumonia. He was taken from the train in Harbin, Manchuria (now Haerpin in Heilungkiang province) to the German-Russian hospital there. When summoned to Marbut's bedside, Thorp recalled Marbut telling him, as he neared death at age 72, "You've got to get me out of here; I have ten more years of work to do." Unfortunately, he died early the next morning. According to Thorp, the American Consulate held a beautiful memorial service at which a Russian refugee chorus sang. His ashes were sent back to southwestern Missouri.

In Service to Humanity

Marbut made his own independent place in the emerging discipline of soil science, but his concerns became more than the scientific facts of soils and their classification. His subjects at professional meetings and in his publications became more international in scope. His presidential address to the Association of American Geographers in December 1924 dealt with the Malthusian theory. He also examined Soviet soils and the capacity for that country's wheat production potentially to rival that of the USA.

The AGS, in awarding him the Cullum Medal in 1932, its highest distinction, described his achievements in the following words: "For geographical works on the study of soils – the foothold of all things." He also received awards from Berlin and Czechoslovakia, and honorary doctorates from UMC in 1916, and Rutgers in 1930. About this and other honors he received, Marbut was grateful but, as he indicated in his correspondence to Louise, they were but a means to an end. He wrote Louise that all his recognition and achievements were important only as they could be used to strengthen his position and influence to bring to the countries of the world knowledge of their soil resources and, thereby, self-sufficiency in agricultural production so that food would not be a reason to wage war. He wrote that he believed it was his duty and responsibility as a Unitarian scientist to do all one man could do to ensure that the countries of the world could feed themselves without having to rely on other countries for food. Marbut shared his unspoken goal of knowledge in the service of humanity, a point-of-view advocated by Unitarians such as Shaler and Davis, in his correspondence with Louise. However, she never mentioned his philosophy in the biographical sketch she wrote for the SSSA memorial volume; it remains only within the letters that she transcribed and incorporated into the manuscript biography of her father.

Epilog

Curtis Fletcher Marbut is mentioned along with Russian Vasilli V. Dokuchaev and American Eugene W. Hilgard when speaking of the 'founders' of modern soil science. He was a pivotal person – well-known as a teacher and promulgator of ideas. Marbut had to battle provincialism to achieve acceptance of the Russian ideas in the USA that he believed necessary to the advancement of the science. He was the world authority on soils in the 1920s and 1930s. Marbut became a scientist in demand worldwide for his knowledge, a circumstance that prompted US President Franklin D. Roosevelt to grant Marbut three yearly extensions beyond the mandatory retirement age of 70. As Thorp said in his eulogy: "[W]ho of us can fail to aspire to a long life such as his, with its earlier hardships, struggles and disappointment and its ultimate triumph; and who can but wish to follow our chosen work until we are old . . . His life was full and rich and he passed on his flaming torch to younger hands while he was still running a strong race." Marbut was installed in the US National Agricultural Hall of Fame in 1989.

Further Reading

Chorley RJ, Beckinsdale RF, and Dunn AJ (1973) *The History of the Study of Landford and the Development of Geomorphology. Volume of the Life and Work of William Morris Davis*. London: Methuen.

Helms D, Effland ABW, and Duran PJ (eds) (2002) *Profiles in the History of the US Soil Survey*, pp. 43–46. Ames, Iowa: Iowa State Press.

Krusekopf HH (ed.) (1942) *Life and Work of C.F. Marbut/Soil Scientist/A Memorial Volume Published by the Soil Science Society of America*. Columbia, Missouri: The Artcraft Press.

Livingstone DN (1987) *Nathaniel Southgate Shaler and the Culture of American Science*. Tuscaloosa: University of Alabama Press.

Marbut CF (1896) Physical features of Missouri. *Missouri Geological Survey (1st series)* 8: 227–316.

Marbut CF (1910) Soils of the Ozark region. *Missouri Agricultural Experiment Station Bulletin*, no. 3.

Marbut CF, translator (1927) *The Great Groups of Soil and their Development by K.D. Glinka*. Ann Arbor, Michigan: Edwards Brothers.

Marbut CF (1928) *Soils, Their Genesis and Classification and Development/A Course of [30] Lectures given in the Graduate School of the United States Department of Agriculture/February to May, 1928*. Mimeograph.

Washington, DC: [USDA], The first 10 lectures of this series were published in 1951 (see below).

Marbut CF (1935) *Soils of the United States. USDA Atlas of American Agriculture*, Part 3. Washington, DC: US Government Printing Office.

Marbut CF (1951) *Soils: Their Genesis and Classification/A Memorial Volume of Lectures Given in the Graduate School of the United States Department of Agriculture in 1928*. Madison, Wisconsin: Soil Science Society of America.

Marbut CF and Schantz HL (1923) *The Vegetation and Soils of Africa*. Research Series No. 13. Washington, DC: American Geographical Society and National Research Council.

Morrow L (1981) An Ozarks landmark: the Curtis Fletcher Marbut house. *White River Valley Historical Quarterly* 7(9): 6–9.

Morrow L (1999) Curtis Fletcher Marbut. In: Christensen LO, Foley WO, Kremer GR, and Winn KH (eds) *Dictionary of Missouri Biography*, pp. 518–519. Columbia: University of Missouri Press.

Tandarich JP (1998) Pedology – disciplinary history. In: Good G (ed.) *Sciences of the Earth: An Encyclopedia of Events, People, and Phenomena*. vol. 2, H–Z, pp. 666–670. New York: Garland Publishing.

Tandarich JP (2002) The development of soil science in the western world to 1927. In: Lal R (ed.) *Encyclopedia of Soil Science*, pp. 659–662. New York: Marcel Dekker.

Tandarich JP and Johannsen CJ (1985) Introduction to the life and work of Dr Curtis Fletcher Marbut. *Soil Survey Horizons* 26(1).

Tandarich JP and Sprecher SW (1994) The intellectual background for the factors of soil formation. In: Amundson R, Hardin J, and Singer M (eds) *Factors of Soil Formation: A Retrospective*, pp. 1–13. Special Publication 33. Madison, Wisconsin: Soil Science Society of America.

Tandarich JP, Johannsen CJ, and Wildman WE (1985) James Thorp talks about soil survey, C.F. Marbut, and China. *Soil Survey Horizons* 26(2): 5–12.

Tandarich JP, Darmody RG, Follmer LR, and Johnson DL (2002) The historical development of soil and weathering profile concepts from Europe to the United States of America. *Soil Science Society of America Journal* 66(2): 335–347.

Thorp J (1935) *Curtis Fletcher Marbut passes on*. Peiping, China: privately printed.

Matric Potential *See* **Hydrodynamics in Soils**; **Water Potential**; **Water Retention and Characteristic Curve**

MEDITERRANEAN SOILS

J Torrent, Universidad de Córdoba, Córdoba, Spain

The term ‘Mediterranean’ is applied to a specific climate and kind of vegetation that is typical of areas surrounding the Mediterranean Sea but also found in five other world regions, namely California, central Chile, Southwest and South Australia, and the Cape region of South Africa. Although there is no consensus on the precise delineation of the Mediterranean region, the total area with a true Mediterranean climate is probably some $2\times10^6\,\mathrm{km}^2$, the Mediterranean region *sensu stricto* occupying approximately $1.2\times10^6\,\mathrm{km}^2$. Because soil-forming factors vary widely among Mediterranean areas, the resulting soils share neither morphology nor genesis. Despite their high diversity, the body of properties of soils in the Mediterranean areas makes them significantly different from soils in subtropical and other temperate climatic regions. This justifies separate consideration of the genesis and properties of ‘Mediterranean soils,’ a term that is used here with no intended specific genetic meaning and encompasses all soils in areas with a Mediterranean climate.

Soil-Forming Factors

Climate

In Köppen’s classification scheme, the Mediterranean climate is designated as a warm, temperate, rainy climate with dry summers (*Cs*), the wettest winter month receiving at least three times as much rain as the driest summer month. Areas with a Mediterranean climate lie mostly around the Mediterranean sea and at latitudes of 30–45° on the west coast of some continents and thus form transitional zones between the influence of the westerlies of higher latitudes and subtropical high-pressure cells. The winter weather is dictated by the fronts of the westerlies, whereas the dry summer reflects the dominance of the midtropospheric anticyclone.

The Mediterranean climate encompasses an interesting variety of subclimates, as illustrated by the temperature and rainfall data from selected stations listed in Table 1. The mean annual precipitation ranges from 300 to more than 1500 mm, with one or two maxima in the rainy season. Summer drought lasts from less than 2 to approximately 5 months. The mean annual temperature ranges from about 11°C (e.g., in regions of Anatolia and central Spain) to approximately 19°C (e.g., in coastal areas of Cyprus), and absolute maxima above 45°C and minima below −20°C have been recorded at some stations. The difference between the mean temperatures of the warmest and coldest month is less than 8°C in some coastal areas affected by cold ocean currents (e.g., San Francisco Bay, California) but exceeds 20°C in inland areas with little oceanic influence. The difference at stations by the Mediterranean Sea is typically 13–16°C. From the standpoint of pedoclimate, the contrast in moisture content between winter and summer was used as the basis for defining the *xeric* moisture regime, an important diagnostic criterion in the soil taxonomy classification system. Under this regime, the soil zone affecting plant growth is dry for more than 45 days in summer and moist for more than 45 days in winter.

The influence of the Mediterranean climate on soil development can be gleaned by considering two simple parameters, namely: the degree of leaching (L) and the actual evapotranspiration (AET). The annual value of L is calculated as the sum of the excess precipitation (P) over potential evapotranspiration (PET) for those periods in which $P > PET$ less the soil water-holding capacity (WHC). The annual value of AET is equivalent to $P - L$. Table 2 shows the water balance for a station with a warm continental Mediterranean climate (Córdoba, Spain) and includes the mean monthly and annual values of P, PET, AET, and L. Winter leaching (37 mm) is low, consistent with other Mediterranean stations with mean annual P below 600–700 mm, if mean monthly P data are used in the calculations. As can be seen from Table 2, the summer drought leaves the soil with little water to meet plant needs and enable mineral weathering and biological activity. The intensity of these processes peaks in autumn and late spring when soil is moist and temperatures are mild. In summary, the relative lack of coincidence between high soil temperatures and high moisture contents precludes strong weathering and continuous biological activity.

Variability in annual rainfall is high in the Mediterranean region. For instance, the annual rainfall in Córdoba for the 1901–2000 period ranged from less than 250 mm to more than 1250 mm, i.e., by more than five times. The annual leaching for the same period ranged from 0 to more than 650 mm (at WHC 125 mm), with a moderately bimodal distribution (Figure 1). Because of the high variability in rainfall, L takes relatively large values in some years. Mean climatic data thus lead to gross underestimation of the leaching potential of the Mediterranean

Table 1 Climatic data from selected stations in the Mediterranean region

Month *Station*	*J*	*F*	*M*	*A*	*M*	*J*	*J*	*A*	*S*	*O*	*N*	*D*	*Year*
Athens													
Temperature (°C)	9	10	11	15	20	24	27	27	23	19	15	11	18
Precipitation (mm)	62	36	38	23	23	14	6	7	15	51	56	71	402
Rome													
Temperature (°C)	8	9	11	14	18	22	24	24	22	17	13	10	16
Precipitation (mm)	83	73	52	50	48	18	9	18	70	110	113	105	749
Cuenca (Spain)													
Temperature (°C)	3	4	7	10	13	18	22	21	18	12	7	4	12
Precipitation (mm)	43	41	70	48	72	50	19	26	43	55	49	55	571
San Francisco													
Temperature (°C)	9	11	12	13	15	16	17	17	18	16	13	10	14
Precipitation (mm)	102	88	68	33	12	3	0	1	5	19	40	104	475
Perth (Australia)													
Temperature (°C)	23	24	22	19	16	14	13	14	15	16	19	22	18
Precipitation (mm)	7	12	22	52	125	192	183	135	69	54	23	15	889

Mean monthly and annual temperature and precipitation values are rounded to the next unit.

Table 2 Soil water balance (mm) for Córdoba (southern Spain; 90 m, 38°51′ N, 4°50′ W), on the basis of the agricultural year, September–August (mean of 100 years: 1901–2000 period)

Month *Variable*	*S*	*O*	*N*	*D*	*J*	*F*	*M*	*A*	*M*	*J*	*J*	*A*	*Year*
P	27	74	91	87	73	73	79	61	45	19	3	3	635
PET	134	78	46	33	38	48	75	108	140	178	218	194	1288
P − *PET*	−106	−4	45	54	35	25	4	−47	−95	−159	−215	−191	−653
Soil storage[a]			45	99	125	125	125	78					
L					8	25	4						
AET	27	75	46	33	38	48	75	108	123	19	3	3	598

[a] For a soil with a water-holding capacity of 125 mm.

P, mean precipitation; *PET*, potential evapotranspiration according to Penman; *L*, leaching; *AET*, actual evapotranspiration.

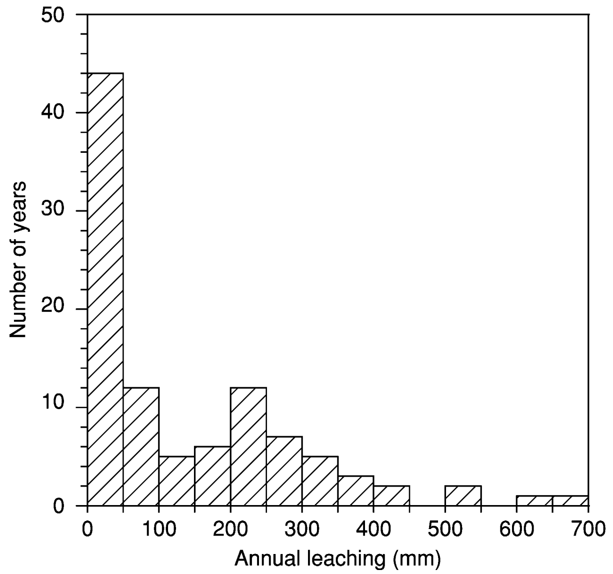


Figure 1 Frequency distribution of annual leaching in Córdoba, southern Spain, for the 1901–2000 period, as calculated from monthly rainfall data for the period, the mean monthly potential evapotranspiration (Table 2), and a soil water-holding capacity of 125 mm.

climate, which is generally significant if the mean annual *P* exceeds 500 mm.

Vegetation

Most Mediterranean areas were once covered with a sclerophyllous forest, which adapts itself readily to both summer drought and light winter frost. The exact nature of this forest in the areas bordering the Mediterranean Sea is difficult to ascertain, but the native vegetation probably included coniferous and broad-leaved evergreen trees. Among the latter, holm oak (*Quercus ilex* and *Q. rotundifolia*), cork oak (*Q. suber*), and Kermes oak (*Q. coccifera*) are widespread. Typically sclerophyllous species include wild olive (*Olea europaea*), carob (*Ceratonia siliqua*), and lentisk (*Pistacia lentiscus*). Isolated specimens of stone pine (*Pinus pinea*) and masses of Aleppo pine (*P. halepensis*) are also commonplace. Degradation of the original forest has often resulted in the establishment of scrub communities that are given local names (*garriga* in Catalonia, *maquis* in France, *macchia* in Italy, etc.) and include cistus (*Cistus* spp.), heathers (*Erica* spp.), gorse (*Ulex* spp.), broom (*Genista* spp.), and thyme (*Thymus* spp.). Some scrub communities probably constitute the original vegetation in areas where either low precipitation or pervious rocks and wind exposure gave rise to habitats that were too dry for the sclerophyllous forest to survive.

In North America, the sclerophyllous forests of western California are rich in species of *Quercus* and *Cupressus* similar to those of the Mediterranean basin (e.g., *Q. agrifolia* resembles *Q. ilex*). In areas with mean annual rainfall below 500 mm, a formation similar to the maquis, namely the *chaparral*, reaches 3 m in height and encompasses bush oaks in addition to species of *Ceanothus* and *Arctostaphylos*. In South America, the Chilean sclerophyllous forest exhibits a completely different floristic composition but is remarkably similar in appearance to analogous formations; *Quillaja saponaria*, *Rhus caustica*, and *Peumus boldus* are especially commonplace. Uncultivated land in the Cape region is now occupied by sclerophyllous scrub called *fynbos*, which varies between 1 and 4 m in height and exhibits a bluish-green hue; the only tree is *Leucodendron argenteum*, which is accompanied by scrub species pertaining to the Proteaceae, Brassicaceae, and Rosaceae families. Finally, in the Southwest and South Australia Mediterranean areas, the vegetation typically consists of *Eucalyptus* species with coriaceous leaves. The *jarrah* forest with *E. marginata* is typical of areas with a mean annual rainfall of 600–1200 mm; on the other hand, the drier (500–600 mm) *wandoo* zone has a less dense forest of *E. redunca*.

Sclerophyllous vegetation seems to be efficient in cycling of bases; however, relatively little information is available on the biogeochemical cycles for the different areas. Transformation of fresh organic matter usually results in a mull humus that grades into a moder humus in the subhumid borders of the Mediterranean region, particularly when soils develop on coarse-grained, acidic parent materials.

Lithology and Relief

A number of Mediterranean areas are mountainous and exhibit a complex lithology that derives from the strong tectonic activity typical of zones involved in crustal plate collision (Mediterranean Sea, California, central Chile). Sedimentary rocks predominate, but plutonic, volcanic, and metamorphic rocks are well represented in most areas. In the Mediterranean region proper, limestones and dolomites constitute a fairly high proportion of the mountains. Frequent erosion and deposition episodes since the Upper Tertiary have led to the formation of large surfaces covered by fine-grained sediments at the foot of the mountains.

The abundance of steep slopes has contributed to make water erosion the principal process in slope development, which was particularly active in Quaternary periods with scarce or uneven rainfall and soils without a continuous vegetation cover. In periods of more-even rainfall, slope development under a continuous vegetation cover was probably more markedly influenced by soil creep and subsurface water erosion, similarly to slopes in humid temperate regions. The strong influence of water erosion is apparent in the slope profiles of many lithologically uniform areas, which, as predicted by geomorphic models, consist of a long, concave element at the foot, a straight middle segment, and a short, convex summit.

Structural reliefs with parallel-retreating slopes are commonplace in areas of sedimentary rocks (particularly in the drier regions). These slopes commonly merge into pediments. Depositional pediment surfaces, which are more common in Mediterranean areas than in humid temperate regions, generally have slopes of less than 1–2% and occasionally extend over tens of square kilometers. Some such surfaces are very old and covered by strongly weathered soils. Not rarely, opposite pediment surfaces form V-shaped, wide valleys dissected to a greater or less extent by a central stream and its tributaries.

Alluvial plains and their contiguous terraces dominate landscapes over stretches of tens of kilometers in the middle and lower parts of major river valleys, where they support irrigated agriculture. River terrace systems provide excellent soil chronosequences for the study of the major soil-forming processes as the systems formed over long periods during the Quaternary. For instance, the oldest river terraces reach heights of more than 200 m above the present river channel in the river Guadalquivir valley (southern Spain).

Time

Because the Mediterranean region underwent no glaciation, its landscape exhibits geomorphic surfaces of widely different age. Some surfaces (e.g., the *raña* surfaces in several parts of Spain) date back to the Pliocene. Lower Pleistocene surfaces are not uncommon, and are well represented by different pediments and river terraces. An intricate mosaic of soil formation times is therefore the rule rather than the exception in most areas. Unfortunately, absolute dating is impossible in most cases, so many pedologists rely on geomorphic data to establish relative soil age and its relationship with soil development.

Soil-Forming Processes

Mass Additions and Losses

The strong influence of water erosion results in marked catenary differentiation in mountainous and hilly areas. Thin, occasionally skeletal soils in the upper part of the slope give way to increasingly thick soils as the footslope is approached. Increasing thickness is usually associated with increasing textural fineness and with changes in the mineralogical assemblage of the clay fraction.

Some Mediterranean areas lie at the fringe of deserts (e.g., the Sahara and Negev deserts). This has resulted in more or less frequent episodes of eolian dust addition. Direct measurement of the dust carried by ‘mud rains’ and detailed mass balances in soils close to the Mediterranean have yielded or supported figures of a few grams per square meter and per year. Dust accretions, gradually assimilated by the soil during the Quaternary, may account for up to 50% of the mass of some soils. Unfortunately, balance calculations are difficult because the receiving soils and the dust contain similar minerals.

Leaching and Redistribution of Calcite and Silica

Weak-to-moderate winter leaching enables weathering of silicates and significant translocation of basic cations. Thus, mature soils that developed on base-poor parent materials exhibit acid pH values in the surface horizons, with pH and base saturation generally increasing with depth. If the parent material contains calcite, dolomite, or Ca-rich weatherable minerals and winter leaching is scarce to moderate, Ca precipitates as secondary calcium carbonate (in the form of calcite) in some soil horizons. Calcic horizons with carbonate nodules of different forms and sizes occur in soils in areas with $P < 700$–800 mm

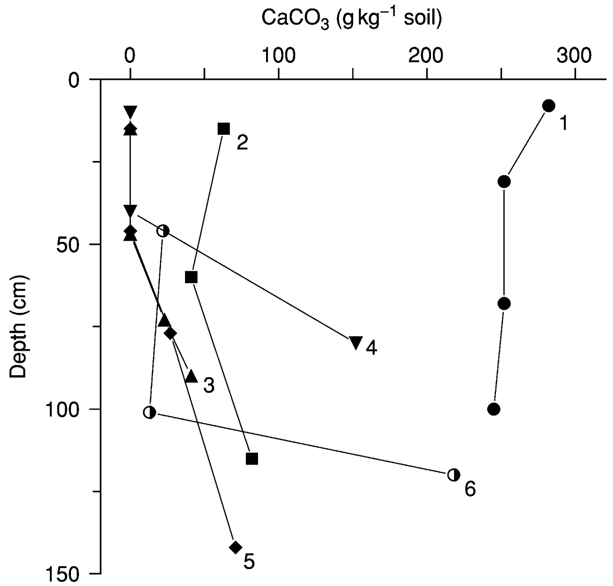


Figure 2 Depth distribution of CaCO$_3$ in soils of a river terrace sequence (Henares River, central Spain). Numbers denote relative soil age (1, recent alluvium; 6, highest terrace). Data from Daz MC (1986) *Estudio de los Procesos de Rubefacción de los Suelos de las Terrazas de la Cuenca del Ro Henares y del Ro Jarama*. PhD thesis, Universidad Politécnica de Madrid, Spain.

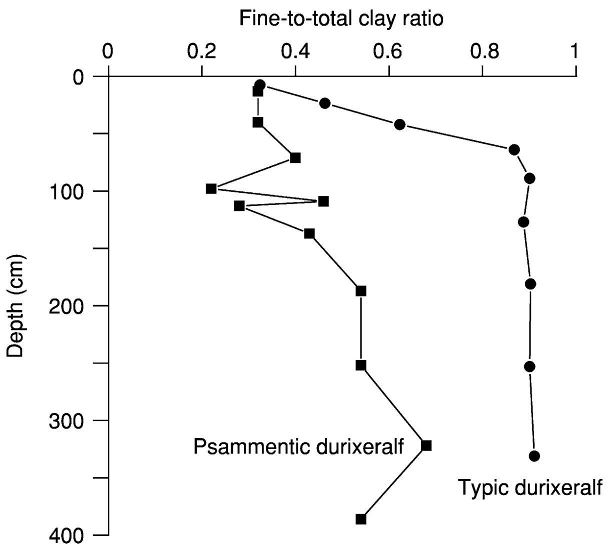


Figure 3 The increase with depth in the fine-to-total clay ratio in two deep soils of southern California is taken as evidence for clay translocation. Data from: Torrent J, Nettleton WD, and Borst G (1980a) Clay illuviation and lamellae formation in a Psammentic Haploxeralf of Southern California. *Soil Science Society of America Journal* 44: 363–369; and Torrent J, Nettleton WD, and Borst G (1980b) Genesis of a Typic Durixeralf of southern California. *Soil Science Society of America Journal* 44: 575–582.

and are replaced by indurated calcrete (petrocalcic) horizons where $P < 400$–500 mm. Long stretches of the dry Mediterranean areas (e.g., La Mancha region in central Spain) are covered by soils with calcrete. The depth at which carbonate accumulates is a function of not only the climate and parent material, but also relief and, particularly, time (Figure 2).

The weathering of silicates releases Si into solution; dissolved silica is translocated and can precipitate as silica (opal) in the form of nodules or crusts (duripans). This process is generally restricted to the drier areas, provided the parent material contains highly weatherable silicates or volcanic glass. Therefore, silica and mixed silica–carbonate are less abundant than carbonate accumulations in Mediterranean soils.

Clay Translocation

Clay translocation (illuviation) has been, and continues to be, a major process in the genesis of Mediterranean soils, particularly in the more humid areas and periods. Because of the strong flocculating effect of Ca ions, decalcification is required for clay particles to disperse and be carried by percolating water; however, translocation of clay has occasionally been observed in calcareous soils. Summer drought seems to favor clay translocation, probably because, when the soil dries, cracks develop and clay on the freshly exposed surfaces is easily dispersed and transported by the water percolating after the first autumn rains. Clay translocation is generally a slow process, but soils with horizons rich in illuviated clay (argillic horizons) have formed within only 2000–3000 years on the alluvium of some rivers of Spain. Summer drought generally contributes to disruption of the clay cutans resulting from clay translocation. Thus, evidence for clay translocation is provided by the presence of undisturbed cutans in deep soil horizons, or by the increase in the fine-to-total clay ratio with increasing depth (Figure 3), since the fine clay particles are assumed to be more mobile than the coarse ones.

Rubefaction

Rubefaction (i.e., reddening due to pedogenic hematite formation) is quite common in the Mediterranean region. It follows decalcification because the high pH of calcareous soils hinders weathering and Fe release from primary minerals. Appropriate environmental conditions for the formation of hematite from ferrihydrite (the initial product of the weathering of Fe minerals) include a high temperature, water activity of less than 1, low organic matter content, near-neutral pH, and high concentrations of Ca^{2+} and Mg^{2+}. In the absence of one or more of these conditions, only goethite (a yellow Fe oxide) is formed.

The vivid red hue of some Mediterranean soils has attracted the attention of pedologists, geologists, naturalists, and travelers. Terms such as the Italian 'Terra Rossa' or 'red Mediterranean soils' have been incorporated into some classification schemes and used (and abused) by pedologists to denote different red-colored Mediterranean soils. Indeed, pedogenic hematite possesses a high pigmenting power, so a concentration of more than than 10 g kg^{-1} soil often results in reddish brown or redder hues. Interestingly, redness ratings derived from the Munsell color notation and the soil hematite content are correlated (Figure 4).

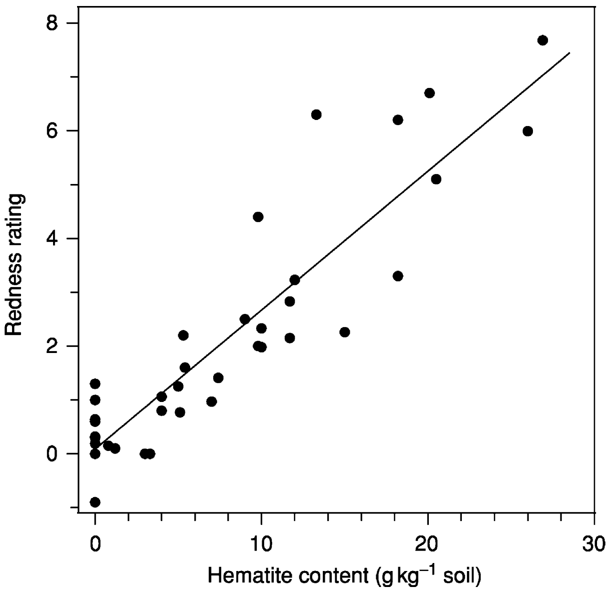


Figure 4 Relationship between a redness rating derived from the Munsell color notation and the hematite content in a group of Mediterranean soils. Data from: Peña F and Torrent J (1984) Relationships between phosphate sorption and iron oxides in Alfisols from a river terrace sequence of Mediterranean Spain. *Geoderma* 33: 283–296; Torrent J and Cabedo A (1986) Sources of iron oxides in reddish brown soil profiles from calcarenites in southern Spain. *Geoderma* 37: 57–66.

Soil chronosequences indicate that rubefaction and clay illuviation develop at similar rates once the upper soil horizons have been decalcified. These three processes play a crucial role in soil development on stable geomorphic surfaces. This is shown in Figure 5, which depicts an idealized development sequence on calcareous alluvium under an 'average' Mediterranean climate. Soils comparable with members of this sequence are seen on widely different parent materials that range from hard igneous rocks to unconsolidated sediments. Two related development sequences must be considered here. One has soils with no carbonate accumulation, so it must be assigned to base-poor rocks. In the other, the old soils exhibit a seasonally perched water table that develops as a result of pore clogging by illuviated clay. In this case, reductomorphic features appear and the red hue is lost due to reductive hematite dissolution.

Development of Vertic Features

The xeric regime favors development of the vertic features typical of Vertisols and other clayey soils that usually experience several major shrink/swell cycles each year. Vertic features seem to develop over short pedogenic times in clayey sediments as recent alluvium or Tertiary marls. Also, vertic features are enhanced by smectite formation in the soils

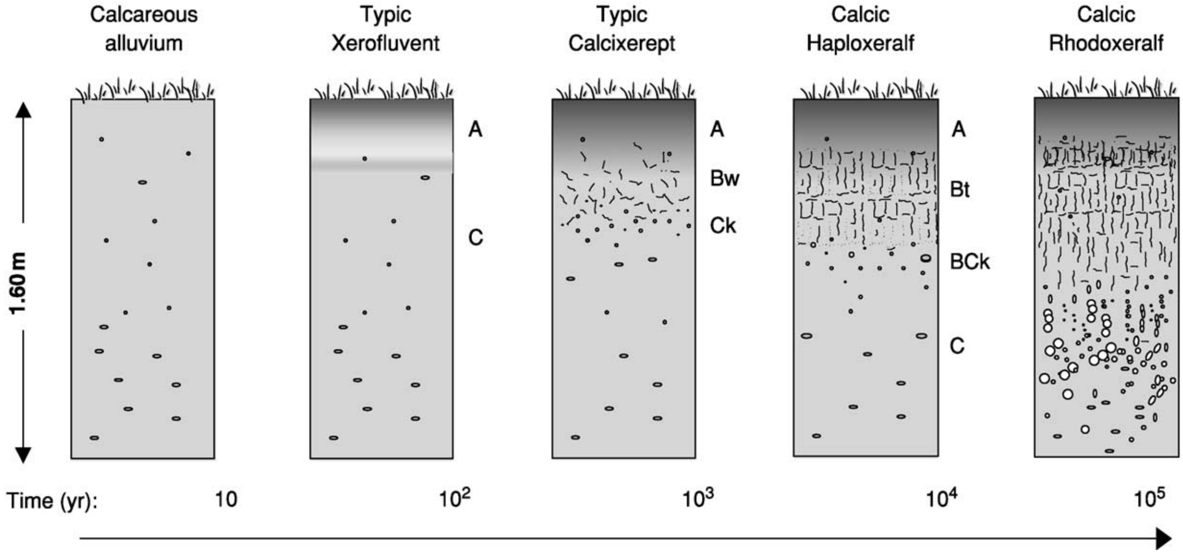


Figure 5 Idealized soil development sequence from a calcareous loamy alluvium in a typical Mediterranean environment.

of footslopes and closed basins as a result of lateral illuviation of silica and bases. For this reason, Vertisols occupy the lowest topographic positions of some Mediterranean soilscapes, even though the parent material may not be clayey.

Major Soil Types

Many pedologists would subscribe to the idea that the Mediterranean climate dictates the predominance of some specific soil-forming processes and thus justifies the term 'Mediterranean soils.' However, pedodiversity is probably higher in the Mediterranean region than in any other climatic zone. How can these two conflicting views be reconciled? Probably by remembering that: (1) the Mediterranean climate encompasses a rich variety of subclimates and (2) the effect of the Mediterranean climatic and biotic environments is modulated by the high lithological and geomorphic diversity resulting from the complex tectonic and climatic history of Mediterranean regions.

Various classification schemes have been used – some only at a national level – to encompass the Mediterranean pedodiversity. Some of the older schemes incorporated geographical and color terms (e.g., 'Black Andalusian Earths') with somewhat diffuse definitions. Most pedologists have by now adopted the UN Food and Agriculture Organization (FAO) legend (or the more recent International Society of Soil Science (ISSS)–International Soil Reference and Information Center (ISRIC)–FAO *World Reference Base for Soil Resources*) and the Soil Taxonomy System developed by the Soil Survey Staff of the United States Department of Agriculture (USDA). The latter is used here for a cursory description of the main soil types.

Eleven of the 12 orders in the soil taxonomy are represented in the Mediterranean region: Alfisols, Entisols, Inceptisols, Mollisols, and Vertisols are commonplace to abundant; Aridisols and Ultisols can be locally important; Andisols, Histosols, and Spodosols are rare; and Oxisols, as those found in some areas of California, are considered to be relict soils.

Alfisols are widespread and correspond to most of the development stages of the chronosequence in Figure 5 (and similar chronosequences). Alfisols are generally found on stable geomorphic surfaces over a wide variety of parent materials and are represented by the Aqualf and Xeralf suborders. The former are locally significant in depressions, closed basins, and plateaus, and they tend to exhibit a perched rather than a permanent water table. Among the Xeralfs, Durixeralfs (with duripan) are generally restricted to volcanic areas, and Fragixeralfs (with fragipan) and Plinthoxeralfs (with plinthite) are relatively rare. Rhodoxeralfs, which exhibit vivid red colors, are commonplace – but not so much so as the Haplo- and Palexeralfs – and cover a variety of well-drained parent materials (calcarenites, hard limestones, arkoses, sandstones, colluvia of acid rocks, basalts, etc.). Terra Rossa *sensu stricto* (i.e., a red, clayey soil on pure, hard limestone) generally belongs to the Rhodoxeralf group. Reddish-brown Xeralfs not meeting the stringent color requirements for Rhodoxeralfs are also commonplace and generally classed as Haploxeralfs. Palexeralfs ('old' Xeralfs) generally occupy old geomorphic surfaces such as high river or marine terraces and ancient pediments. Some possess a petrocalcic horizon; others tend to support a perched water table. In any case, Haploxeralfs, which are regarded by many as the ragbag of Xeralfs, constitute the most widespread great group.

Entisols, which exhibit minimal pedogenic alteration, are particularly abundant in the mountainous or hilly areas, where water erosion and other mass-wasting processes keep the soil permanently thin. These soils are usually classed as Xerorthents. Less abundant, but agriculturally important, are Xerofluvents, which occupy alluvial plains and some colluvial areas. Psamments (sandy Entisols) are confined to alluvia, coastal dunes, and sandy continental deposits, as are Aquents (wet Entisols) to some depressions, deltas, and alluvial flats. Not uncommonly, the soils of these two suborders underlie highly ecologically significant rather than agriculturally valuable areas.

Inceptisols in Mediterranean areas often represent steps in evolutionary sequences ranging from raw soils (Entisols) to Alfisols or Vertisols, but they are truly mature soils in other cases. Most are Xerepts, but Aquepts are also present. Xerepts include soils with calcic or petrocalcic horizons (Calcixerepts) and acidic soils (Dystroxerepts). They occur in all types of landscape and climatic area, and include agricultural and forest soils. Aquepts are restricted to alluvial and coastal plains, and closed basins; also, like Aquents, some are environmentally significant and can be turned into good agricultural soils after drainage.

The presence of a dark, well-structured (mollic) epipedon (a requirement for the Mollisol order) was probably common at some time under the typical sclerophyllous vegetation. Man's activities (deforestation, ploughing, grazing, etc.) have resulted in substantial losses of organic matter and deterioration of the good structural properties of the epipedon. Assessing the extent of such changes, however, is a difficult task. At present, less than 10% of the Mediterranean region is covered by Mollisols (mostly Xerolls), which tend to be concentrated in mountainous areas, man-made grasslands, and other uncultivated areas.

Table 3 Soil types most commonly found in Mediterranean areas

Order[a]	*Suborder*[a]	*Group*[a]	*WRB*[b]	*Probable extent (%)*
Entisols	Orthents	Xerorthents	Leptosols, Regosols	20–25
	Fluvents	Xerofluvents	Fluvisols	1–3
	Psamments	Quartzipsamments	Arenosols	<1
		Xeropsamments	Arenosols	1–3
Inceptisols	Xerepts	Calcixerepts	Calcisols	10–15
		Dystroxerepts	Cambisols	3–5
		Haploxerepts	Cambisols	10–15
	Aquepts		Gleysols	<1
Vertisols	Xererts	Calcixererts	Vertisols	1–2
		Haploxererts	Vertisols	1–2
	Aquerts		Vertisols	<1
Alfisols	Xeralfs	Haploxeralfs	Luvisols	15–25
		Palexeralfs	Luvisols	3–5
		Rhodoxeralfs	Luvisols	3–5
		Durixeralfs	Durisols	<1
	Aqualfs		Luvisols	<1
Mollisols	Xerolls	Calcixerolls	Kastanozems	2–4
		Haploxerolls	Phaeozems	2–4
Ultisols	Xerults	Haploxerults	Acrisols	1–2
		Palexerults	Acrisols	<1
Aridisols	Salids		Solonchaks	<1

[a] According to Soil Survey Staff (1999) *Soil Taxonomy*, 2nd edn. Agricultural Handbook 436. Washington, DC: USDA.
[b] ISSS–ISRIC–FAO (1998) *World Reference Base for Soil Resources*. Leuven, Belgium: Acco Press.

Vertisols form readily where the parent material is clayey or the local conditions favor the formation of smectite. Mediterranean Vertisols (Xererts) occupy relatively large areas on Tertiary marls and Quaternary clays (e.g., in the river Guadalquivir valley in Spain and some plains of western Morocco, where smectite is largely inherited from the parent material ('lithogenic' Vertisols)). In closed basins, footslopes, and some poorly drained areas, where smectite is neoformed ('topogenic' Vertisols), Xererts alternate with Aquerts and gradually merge into Alfisols at some higher topographic locations. Among Xererts, Calcixererts are probably more abundant than Haplo- and Durixererts, because the low hydraulic conductivity hinders decalcification in Vertisols.

Ultisols exhibit low base saturation in the subsoil, so they cannot be expected to form under the leaching regime imposed by the average Mediterranean climate. However, they do occur in areas where at least two of the following conditions are met, namely: (1) high (more than 200–300 mm) winter leaching; (2) base-poor parent materials (e.g., slates or quartzites); and (3) a stable geomorphic surface. These three conditions are fulfilled, for instance, by Xerults and Aquults in Pliocene *raña* formations, which occupy more than 10 000 km^2 in Spain.

The only Aridisols present in the Mediterranean regions are Salids. Although not abundant, these saline soils fairly frequently host ecologically unique areas in deltas, enclosed depressions, and coastal plains.

Finally, only the coincidence of unusual local conditions favors the occurrence of Spodosols, Andisols, and Histosols. For instance, Spodosols in western Portugal form on coastal dunes under a pine forest in a subhumid Mediterranean environment.

Table 3 shows the probable extent for the most common soils in the Mediterranean region. These figures can only be tentative, given the relative scarcity of detailed soil maps for some Mediterranean countries.

Physical, Chemical, and Mineralogical Properties

Many lands bordering the Mediterranean have been deforested and cultivated for centuries and even millennia. Man-induced water erosion and organic matter depletion have affected extensive areas, which have been worn down to unacceptable limits. Soil degradation in other Mediterranean areas has been less marked, but current risks derived from agriculture, forestry, industrial activities, or modern urban needs run high in most areas. Soil properties must be envisaged in any case from the standpoint of their significance and evolution in different man-influenced scenarios, which change – sometimes dramatically – with time.

Physical Properties

Generally, the quality of Mediterranean soils is more strongly influenced by physical than by chemical properties. Moreover, some physical properties cannot be significantly modified (e.g., texture) or changed without much effort (e.g., structural stability).

Mediterranean soils tend to be more clayey than soils from cool temperate areas. Textural contrasts such as those exhibited by Alfisols are common and generally result in water-holding capacity increasing with increase in depth, which favors growth of rainfed crops. Stoniness is widespread and generally reduces crop production; however, surface stoniness helps reduce surface evaporation, protects soil against splash erosion, and delays runoff.

The low structural stability of many cultivated soils has been attributed to their low content in organic matter. Various studies, however, have revealed other factors (Fe oxides, Ca saturation, clay type) to be equally important with a view to creating structure. Soil crusting affects many soils rich in silt and fine sand, where it hinders emergence of small seedlings and triggers runoff. Hardsetting A horizons, which combine structural instability and high strength when the soil is dry, have been reported in the Xeralfs of Southwest Australia and some countries bordering the Mediterranean.

Low water availability in summer limits the yield potential of Mediterranean soils, even though they possess, on average, a water-holding capacity exceeding that of soils in humid temperate regions. Soil crusting and low infiltrability may offset the latter advantage, however.

Mineralogical Properties

The lack of coincidence of high temperatures and high moisture contents prevents intensive weathering in the Mediterranean environment. Therefore, highly weathered soils occur only on old, stable geomorphic surfaces in subhumid regions. Generally, because of the low-to-moderate degree of weathering, much of the soil silicate clays is inherited from the parent material or produced by transformation of primary micas rather than neoformed from solution. Illite constitutes approximately 50% of the clay fraction on average, as it abounds in sedimentary rocks and is produced by transformation of muscovite.

The overall sequence of abundance of clay minerals is illite > smectite, illite–smectite > kaolinite > chlorite, vermiculite, hydroxy-interlayered vermiculite >> halloysite, palygorskite, pyrophyllite, talc. The composition of the soil solution in many soils falls near the lines bounding the kaolinite and smectite stability fields. There is indeed evidence that kaolinite in the more leached soil environments and smectite in solute-rich environments are neoformed in substantial amounts. Neoformation of illite has been documented in K-rich solutions, such as those in microcracks of potassium feldspar grains. The presence of different microenvironments, leaching gradients in the soil profile, and lateral eluviation combine to create differences in clay mineralogy with depth and along a catena. Figure 6 depicts an idealized spatial distribution of the minerals present in the clay fraction in a typical catena.

Iron oxides are the most abundant among nonsilicate clay minerals; they typically account for 5–10%

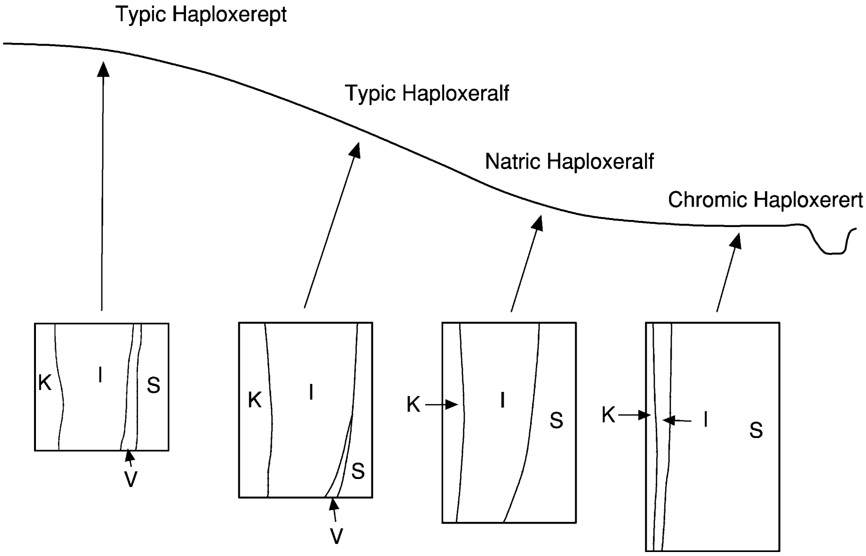


Figure 6 Relative abundances of silicate clays in the clay fraction of soils of a hypothetical catena on a granodiorite saprolite in a typical Mediterranean environment. K, kaolinite; I, illite; V, vermicullite; S, smectite.

of the clay fraction. The Mediterranean pedoenvironment favors crystallization of ferrihydrite to goethite and hematite, so that the ratio of poorly crystalline to crystalline Fe oxides is generally low. Aluminum oxides are rare, gibbsite being present in some excessively drained soils formed on coarse-grained, acidic plutonic rocks.

Among the abundant carbonates, calcite and magnesium calcite are either inherited or secondary, whereas dolomite is always inherited. Calcite crystals can be dispersed in the soil matrix or aggregated in nodules, tubules, or crusts. Crystals range in size from clay to fine sand, but the fine silt fraction is generally the most abundant.

Chemical Properties

The pH of Mediterranean soils generally lies within the slightly acidic to moderately alkaline range and, because of the occasionally steep vertical leaching gradient, can increase substantially with depth. Thus, the idealized mature soils of Figure 5 typically have a pH of approx. 6 on the surface and approx. 8 in the horizons where carbonate accumulates. However, ploughing and accretion of dust from nearby or distant sources often result in recalcification of surface horizons. Consequently, the highly acidic soils formed on old geomorphic surfaces on acidic rocks lie usually far from areas with calcareous soils and rocks.

The characteristic mineral assemblages of Mediterranean soils endow them with interesting quantitative adsorption features. Thus, the cation exchange capacity of the clay fraction generally ranges from 40 to 60 $\mathrm{cmol_c\,kg^{-1}}$ as a result of the prevalence of 2:1 clay minerals. Specific ion adsorption by hydroxylated mineral surfaces largely arises from Fe oxides and, to a lesser extent, from the edges of clay minerals and aluminum oxides. Calcite surfaces also play a prominent role in adsorption and heterogeneous precipitation processes. Quantitative relationships between adsorption capacity and simple mineralogical or chemical properties can be established for most of the inorganic adsorbates of interest (e.g., phosphate).

Generally, the adsorptive properties of Mediterranean soils depend strongly on depth as a result of steep gradients in pH, particle size distribution, and mineralogical composition. Consequently, the appropriate ‘filters’ for the different nutrients and pollutants lie at different depths – and make morphological and mineralogical characterization of the soil profile necessary.

From the standpoint of chemical fertility, Mediterranean soils compare favorably with soils in other geographic areas. After N, P is the most limiting nutrient in native soils; fortunately, most soils are acceptably responsive to fertilization with P. Most are well supplied with K, because of the high illite content, and contain adequate levels of Ca and Mg. However, Fe, Cu, Mn, Zn, and B deficiencies frequently arise in some horticultural crops cultivated on calcareous soils.

Salinization, occasionally accompanied by specific ion toxicity problems, has made its way in both old and recently irrigated areas. Even though the principles for good management of irrigation have been known for decades, salinization continues to be a major issue; the scarcity of good irrigation water in most areas, and the use of water-saving irrigation systems, contribute to the problem.

Further Reading

Daz MC (1986) *Estudio de los Procesos de Rubefacción de los Suelos de las Terrazas de la Cuenca del Ro Henares y del Ro Jarama*. PhD thesis, Universidad Politécnica de Madrid, Spain.

ISSS–ISRIC–FAO (1998) *World Reference Base for Soil Resources*. Leuven, Belgium: Acco Press.

Peña F and Torrent J (1984) Relationships between phosphate sorption and iron oxides in Alfisols from a river terrace sequence of Mediterranean Spain. *Geoderma* 33: 283–296.

Roquero C (1979) The potential productivity of Mediterranean soils. In: *Soils in Mediterranean Type Climates and their Yield Potential* (Proceedings of the 14th Colloquium of the International Potash Institute held in Seville/Spain) pp. 21–42. Bern, Switzerland: International Potash Institute.

Soil Survey Staff (1999) *Soil Taxonomy*, 2nd edn. Agriculture Handbook 436. Washington, DC: US Department of Agriculture.

Torrent J (1995) *Genesis and Properties of the Soils of the Mediterranean Regions*. Napoli, Italy: Dipartimento di Scienze Chimico–Agrarie, Università degli Studi di Napoli Federico II.

Torrent J and Cabedo A (1986) Sources of iron oxides in reddish brown soil profiles from calcarenites in southern Spain. *Geoderma* 37: 57–66.

Torrent J, Nettleton WD, and Borst G (1980a) Clay illuviation and lamellae formation in a Psammentic Haploxeralf of southern California. *Soil Science Society of America Journal* 44: 363–369.

Torrent J, Nettleton WD, and Borst G (1980b) Genesis of a Typic Durixeralf of southern California. *Soil Science Society of America Journal* 44: 575–582.

Yaalon DH (1997) Soils in the Mediterranean region: what makes them different? *Catena* 28: 157–169.

METAL OXIDES

A C Scheinost, Swiss Federal Institute of Technology (ETH), Zurich, Switzerland

Introduction

The oxide minerals typically present in soils comprise oxides, hydroxides, oxyhydroxides, and hydrated oxides of Si, Fe, Mn, Al, and Ti. With the exception of the Si oxide quartz and some Ti oxides, which are predominantly inherited from primary (rock) minerals, most oxides form in soil. The metal cations Fe^{2+}, Mn^{2+}, Ti^{4+}, Al^{3+}, and Si^{4+} are released from silicates by weathering. The divalent cations Mn and Fe oxidize and hydrolyze, and precipitate almost exclusively as oxide minerals. Al^{3+} and Si^{4+} have a strong tendency to form secondary aluminosilicate clay minerals; the formation of Al and Si oxides requires special conditions. All oxides have a very low solubility at common soil pH and are therefore enriched during pedogenesis. Highly weathered soils, which have lost a substantial part of their alkalis, alkaline earths, and Si, may contain as much as 50% wt metal oxides.

Oxides of Fe, Mn, and Al may exhibit a high surface area, with reactive surface sites which strongly bind oxyanions and metal cations, thereby impacting the availability and mobility of plant nutrients and toxic metals. Some Mn oxides have a strong oxidizing power and may degrade organic pesticides, decrease or increase the toxicity of heavy metals, and contribute to the formation of soil organic matter. In contrast, sand-sized crystals of the Si oxide quartz are chemically very inert.

The basic structural unit of Fe, Mn, Al, and Ti oxides are cationic metal centers bound to six oxygens. If one drastically reduces the size of the oxygens in relation to the metal cations, this arrangement can be visualized as an octahedron (Figure 1). The octahedra may be linked to each other in three ways, sharing oxygen corners (1 O), sharing edges (2 O), or faces (3 O). Many oxide minerals such as goethite contain only the first two types of connections, while the Fe oxide hematite contains all three. Silicon is bound to four oxygens, forming a tetrahedron. These tetrahedra are connected via corners only. We will use this polyhedral approach in the text and in figures to introduce the mineral species and highlight their major differences and similarities. For each oxide group, we will give an overview including general formation pathways and influence on soil properties. This overview is followed by a description of the minerals, their occurrence in soil, and specific conditions of their formation.

Fe Oxides

Iron is released by weathering of Fe(II)-containing silicates (biotite, pyroxene, amphibole, olivine). After oxidation to Fe^{3+} and hydrolysis, most Fe precipitates as Fe(III) oxides. (Only a small remaining amount of Fe^{2+} becomes a structural part of pedogenic phyllosilicates.) Under oxic conditions, Fe oxides are very insoluble, hence are enriched during pedogenesis. They readily dissolve, however, under reducing soil conditions:

$$2Fe_2O_3 + CH_2O + 8H^+ \Rightarrow 4Fe^{2+} + CO_2 + 5H_2O \qquad [1]$$

This reaction is driven by microorganisms, which metabolize biomass (in eqn [1] represented by CH_2O), and transfer freed electrons to Fe^{3+}, thereby

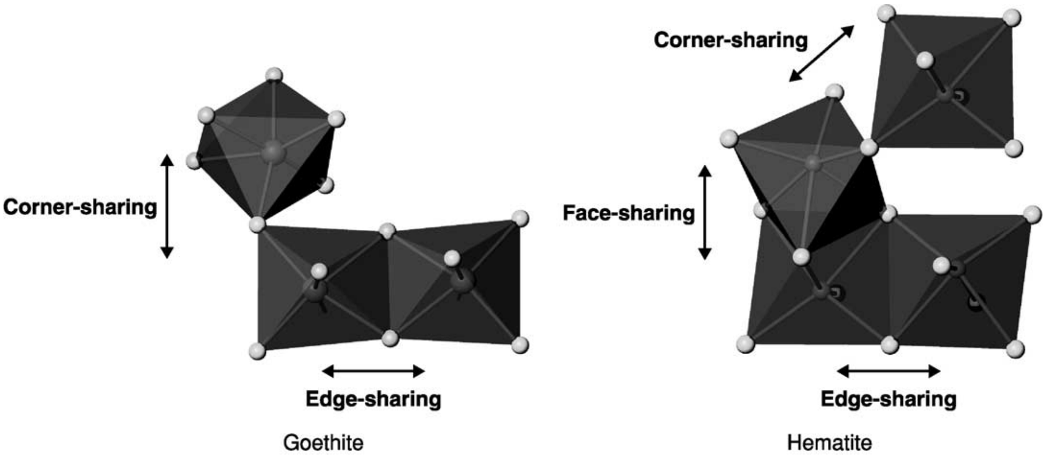


Figure 1 Linkage of metal octahedra by corners, edges, and faces, involving one, two, and three oxygen ligands, respectively. Shown are small structural subunits of the two Fe oxides goethite and hematite.

reducing it to Fe^{2+}. Dissolution of Fe oxides is therefore linked to microbial activity. Depending on parent material, degree of weathering, and soil redox conditions, the amount of Fe oxides in soils varies between less than 1 and more than $500\,g\,kg^{-1}$. The Fe oxides are commonly the strongest pigments in soil, and small quantities may impart vivid colors of red, orange, yellow, brown, and even blue-green (Table 1). Depending on pedogenesis, they occur either evenly distributed in the soil matrix, or locally enriched as skins, patches, and concretions. Soil horizons indurated by Fe oxides form ferricretes or laterites.

Soil Fe oxides commonly exist as very small crystals (2–50 nm), responsible for a large surface area ($50–450\,m^2\,g^{-1}$). Due to the their circumneutral point of zero charge (PZC), they contribute to the anion exchange properties of soils. In addition, their surfaces have a strong affinity for oxyanions and many metal cations. Finally, other metal cations with an ionic diameter similar to Fe^{3+} (e.g., Al^{3+}, Ni^{2+}, Ti^{4+}, Mn^{4+}, Co^{3+}, Cr^{3+}, Cu^{2+}, Zn^{2+}, V^{3+}) may partially replace Fe in the crystal structure of Fe oxides. Due to these properties, Fe oxides play a significant role in the environmental cycling of plant nutrients such as N and P, of transition metals relevant for metabolic reactions (e.g., Mn, Zn, Co), and of toxic elements such as As and U. The 10 Fe oxide minerals observed in soils or sediments at the Earth's surface and some of their diagnostic criteria are listed in Table 1. Crystal structures are shown in Figure 2. Schematic formation and transformation pathways of the eight most common Fe oxides are shown in Figure 3.

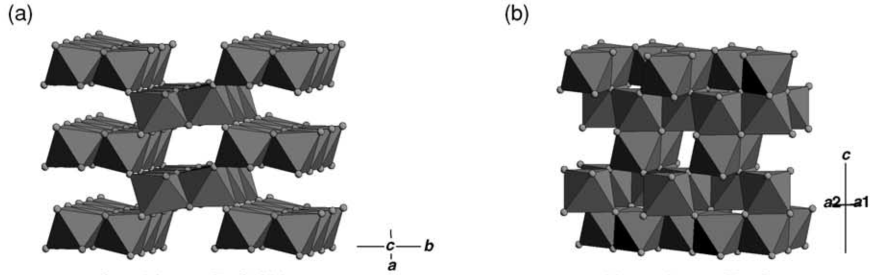
Goethite (α-FeOOH)

Due to its high stability, goethite is the most ubiquitous Fe oxide mineral in soils. In the absence of hematite, its yellowish color is responsible for the yellowish-brown hue of many soils (7.4 YR–3.5 Y). Goethite formation is favored by slow hydrolysis of Fe^{3+} hydroxy cations at low temperature. The even distribution of goethite in a soil horizon indicates therefore that this soil has formed under aerated, temperate, humid conditions. The crystal structure

Table 1 Iron oxide minerals

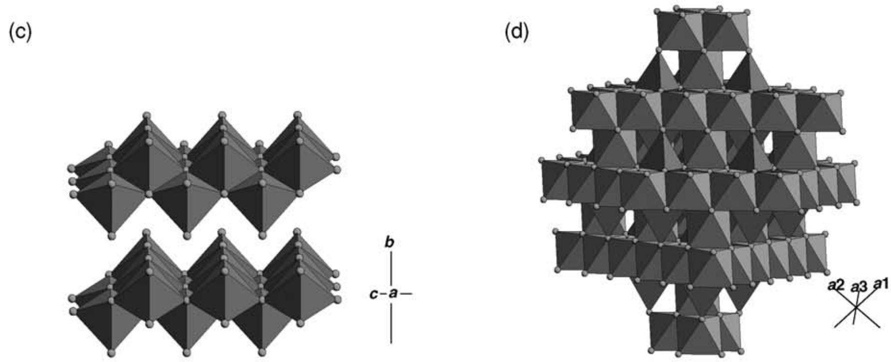
Mineral name and formula	*Structural type*	*Median color*	*Range of hue*	*Typical crystal shape*	*Occurrence*	*Figure*
Goethite[a] α-FeOOH	Diaspore	Strong yellowish-brown	7.3 YR–1.5 Y	Needles, laths	Ubiquitous in soils of all climates	2a
Lepidocrocite[a] γ-FeOOH	Boehmite	Moderate orange	4.9 YR–7.9 YR	Laths	Seasonally anaerobic, noncalcareous soils in cool-temperate climate	2c
Akaganeite β-FeOOHCl	Hollandite	Strong brown	1.2 YR–6.8 YR	Spindle-shaped rods	Hot springs, volcanic deposits	2f
Feroxyhyte δ'-FeOOH		Brownish orange	2.8 YR–9.2 YR	Spheres		2e
Schwertmannite $Fe_8O_8(OH)_6SO_4$	Hollandite	Dark orange yellow	6.2 YR–0.3 Y	Hedgehog aggregates	Acid-sulfate springs and mine drainages	2f
Ferrihydrite[a] $Fe_5HO_8 \cdot 4H_2O$		Strong brown	3.7 YR–5.4 YR	Plates, needles	Frequent in seasonally wet and/or organic-matter rich soils in cool-temperate climate	–
Hematite[a] α-Fe_2O_3	Corundum	Moderate reddish brown	3.5 R–4.1 YR	Hexagonal plates	Frequent in soils of tropical, subtropical, and Mediterranean climate	2b
Magnetite[a] $Fe^{II}Fe_2^{III}O_4$	Inverse spinel	Dark yellowish brown	6 YR–9.5 YR	Laths, cubes	Inherited from magnetite-containing bedrock, microbial origin possible	2d
Maghemite[a] γ-Fe_2O_3	Defect spinel	Black		Octahedra	Common in highly weathered, tropical and subtropical soils	2d
Green rust $Fe^{II}Fe^{III}(OH)_4Cl$	Hydrotalcite (pyroaurite)	Bluish-green		Plates	Permanently wet subsoils	2g

[a] Minerals commonly occurring in soil.



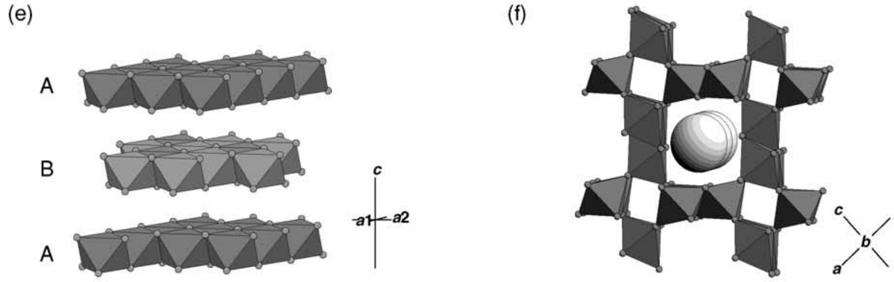
Goethite, α-FeOOH
Ramsdellite, MnO_2
Groutite, MnOOH
Diaspore, α-AlOOH

Hematite, α-Fe_2O_3
Corundum, α-Al_2O_3
Ilmenite, $FeTiO_3$



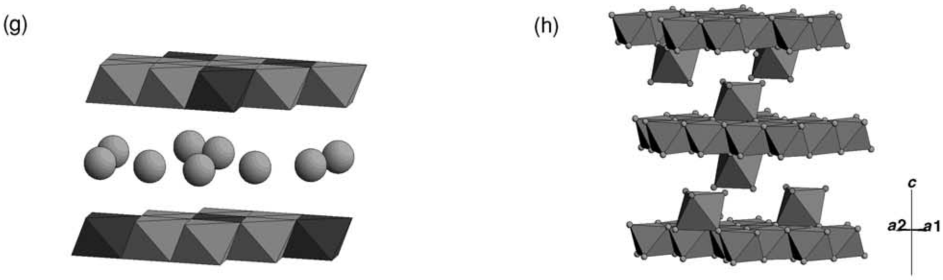
Lepidocrocite, γ-FeOOH
Feitknechtite(β-FeOOH)
Boehmite, γ-AlOOH

Magnetite, Fe_3O_4
Maghemite, γ-Fe_2O_3
Hausmannite, Mn_3O_4
Titanomagnetite, $Fe_{3-x}Ti_xO_4$
Titanomaghemite, $Fe_{2-x}Ti_xO_3$
Ulvöspinel, Fe_2TiO_4



Feroxyhyte, δ'-FeOOH(A=B=Fe)
Buserite(A=B=Mn)
Lithiophorite(A=Mn, B=Al, Li, Zn)
Asbolane(A=Mn, B=Co, Ni)

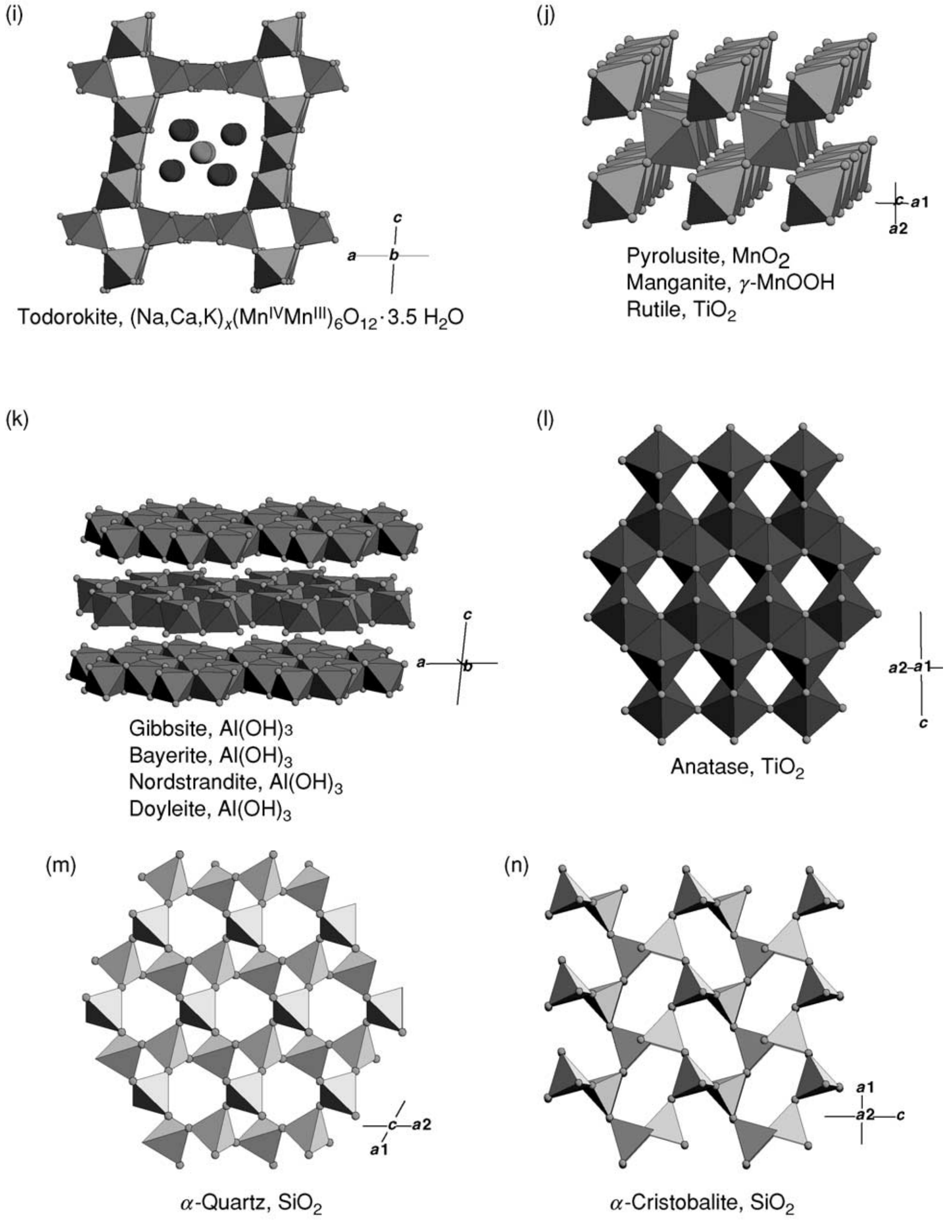
Akaganeite, β-FeOOH
Schwertmannite, $Fe_8O_8(OH)_6(SO_4)$
Hollandite, $Ba_xMn_8O_{16}$
Cryptomelane, $K_xMn_8O_{16}$
Coronadite, $Pb_xMn_8O_{16}$



Green rust, $Fe^{II}Fe^{III}(OH)_4Cl$

Chalcophanite, $ZnMn_3O_7 \cdot 3H_2O$
Birnessite, $(Na,Ca,Mn^{II})Mn_7O_4 2.8H_2O$

Figure 2 Crystal structures of common oxide minerals.



Todorokite, $(Na,Ca,K)_x(Mn^{IV}Mn^{III})_6O_{12} \cdot 3.5\ H_2O$

Pyrolusite, MnO_2
Manganite, γ-MnOOH
Rutile, TiO_2

Gibbsite, $Al(OH)_3$
Bayerite, $Al(OH)_3$
Nordstrandite, $Al(OH)_3$
Doyleite, $Al(OH)_3$

Anatase, TiO_2

α-Quartz, SiO_2

α-Cristobalite, SiO_2

Figure 2 (*Continued*)

of goethite consists of double chains of edge-shared octahedra that are joined to other double chains by sharing corners and by hydrogen bonds (Figure 2a).

Hematite (α-Fe_2O_3)

Soils forming at subtropical or tropical climates often have reddish or even purplish hues (8.0 R–1.4 Y), derived from blood-red hematite. Hematite usually coexists with goethite, but, due to its greater tinting strength, masks the yellowish color of goethite even at low concentrations. Hematite forms by dehydroxylation of ferrihydrite (see below) via a solid-state mechanism which is not yet fully understood. In comparison with goethite, the formation of hematite is favored by neutral pH, increasing soil temperature, and decreasing water activity. The crystal structure of hematite consists of sheets of edge-sharing octahedra. The sheets are connected by edge- and face-sharing octahedra (Figure 2b). The unique face-sharing arrangement and the resulting proximity of neighboring Fe centers is responsible for the red color of this mineral.

Ferrihydrite ($Fe_5HO_8 \cdot 4H_2O$)

This poorly crystalline, hydrous oxide commonly occurs in cooler soils with changing redox conditions. Its formation is favored by rapid oxidation and hydrolysis of Fe^{3+} from solution. The transformation

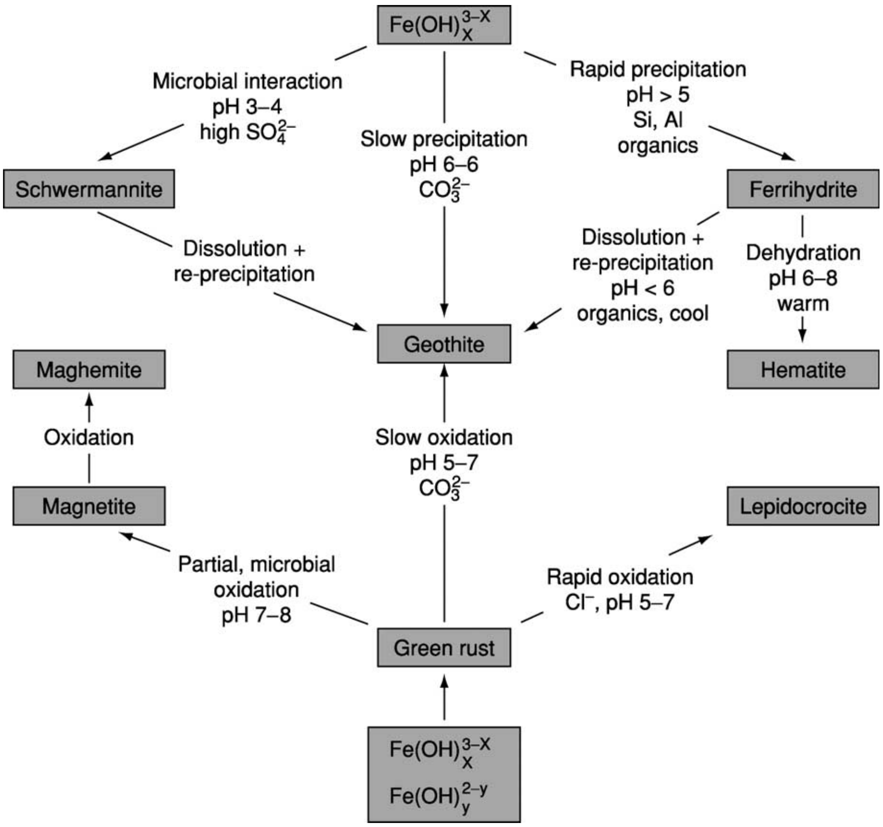


Figure 3 Schematic formation and transformation pathways of common Fe oxides in soils. The precipitation of Fe oxide minerals proceeds either from ferric hydrolysis species (top) or from mixed ferric-ferrous hydrolysis species (bottom). Figure modified after Bigham JM, Fitzpatrick RM, and Schulze DG (2002) Iron oxides. In: Dixon JB and Schulze DG (eds) Soil Mineralogy with Environmental Application, pp. 323–366. Madison, WI: Soils Science Society of America and Cornell RM and Schwertmann U (2003) The Iron Oxides: Structure, Properties, Reactions, Occurence and Uses, 2nd edn. Weinheim, Germany: VCH Verlagsgesellschaft.

into thermodynamically more stable goethite may be hindered by adsorption of silicate, phosphate, and organic anions, making ferrihydrite a common constituent of younger soils in association with goethite. The brownish-orange color of ferrihydrite often marks gradients between anoxic and oxic soil regions, e.g., at the transition between water-saturated and aerated soil horizons, and along root channels or other macropores. Ferrihydrite is also responsible for the brownish-orange color in podzol B horizons. The structure is still being investigated, but can be visualized as a defective hematite structure containing both edge- and face-sharing octahedra.

Lepidocrocite (γ-FeOOH)

This vividly orange, less abundant mineral occurs in seasonally anaerobic, clay-rich, carbonate-free soils in cool climates in association with the more stable goethite. Lepidocrocite forms through relatively slow oxidation of Fe^{2+} at low concentrations of carbonate. Its crystal structure contains double chains of octahedra which are joined by shared edges, resulting in corrugated sheets of octahedra. These corrugated sheets are stacked one on top of the other and are held together by hydrogen bonds (Figure 2c).

Magnetite (Fe_3O_4) and Maghemite (γ-Fe_2O_3)

Blackish magnetite is usually inherited from the parent material, but may also form in soil through biotic processes. The brownish-red maghemite is common in soils of tropical and subtropical climate, but occurs occasionally also in temperate soils that have been exposed to forest or heath fires. Maghemite forms either through oxidation of magnetite, or by heating (300–425°C) of other Fe oxides in the presence of organic compounds. Both magnetite and maghemite have the same inverse spinel structure, with octahedral and mixed tetrahedral/octahedral layers stacked along the direction [111] (Figure 2d). In magnetite, which contains both Fe^{2+} and Fe^{3+} ions, Fe^{3+} occupies both tetrahedral and octahedral sites, while the larger Fe^{2+} ion occupies octahedral sites only.

Rare Fe Oxide Minerals

Several minerals have rarely or never been observed in soils. They may form in extreme natural environments (akaganeite, schwertmannite), in polluted surface waters (schwertmannite, feroxyhite), or transform rapidly into more stable phases (green rust). Feroxyhyte, δ'-FeOOH, has been observed in rusty precipitates of rapidly flowing, Fe^{2+}-containing water, which was quickly oxidized. It consists of sheets of edge-sharing octahedra (Figure 2e). Akaganeite (β-FeOOHCl) has been found in environments with high chloride concentrations ($0.1\,\mathrm{mol\,l^{-1}}$), low pH (3–4), and high temperature (60°C), e.g., in hot springs and volcanic deposits. Akaganeite consists of 2×2 channels (referring to the width of the channels in octahedral units) built by edge-sharing double chains (Figure 2f). These tunnels 0.5 nm in cross section contain Cl^- ions that stabilize the structure. Schwertmannite, $Fe_8O_8(OH)_6SO_4$, is isostructural with akaganeite, but SO_4^{2-} occupies the tunnels (Figure 2f). Schwertmannite has been found in strongly acid sulfate waters associated with mining activities (acid mine drainage) or pyritic rock outcrops. Green rust, $[\mathrm{Fe}^{\mathrm{II}}_{1-x}\mathrm{Fe}^{\mathrm{III}}_{x}(\mathrm{OH})_2]^{x+} \cdot (x/n)\mathrm{A}^{n-} \cdot m\mathrm{H_2O}$ (where A represents anions such as Cl^-, NO_3^-), has been observed rarely in reducing, weakly acid to weakly alkaline subsoils. Its greenish-blue color changes rapidly to yellowish brown on exposure to the air, indicative of its instability against oxidation. Green rust consists of sheets of edge-sharing octahedra, which are occupied by both Fe^{2+} and Fe^{3+}. The resulting positive layer charge is balanced by hydrated anions in the interlayer space (Figure 2g). Due to its anion exchange capacity, redox reactivity and high surface area, even small amounts of green rust may play an important role in aquifers and the vadose zone. Since the mineral forms as a metastable corrosion product of metallic iron at anaerobic conditions, green rust is also relevant for barriers of scrap iron engineered to clean up groundwater.

Mn Oxides

Mn is released by weathering of Mn(II)-containing silicates (biotite, pyroxene, amphibole). After oxidation of the soluble Mn^{2+} to Mn^{3+} and Mn^{4+}, brownish-black Mn oxides of low solubility precipitate, which redissolve only under reducing soil conditions. The reaction equation assuming a Mn(IV) oxide can be written in analogy to the dissolution of Fe oxides

$$2MnO_2 + CH_2O + 4H^+ \Rightarrow 2Mn^{2+} + CO_2 + 3H_2O$$

As for the Fe oxides, the dissolution depends on microbial activity. While oxidation of Mn^{2+} is catalyzed by mineral surfaces, microbial oxidation seems to be the more dominant process in soil environments. Therefore, both formation and dissolution of Mn oxides are intimately linked to microbial activity.

Mn oxides may occur evenly distributed in the soil matrix like the stable Fe oxides goethite and hematite. Due to their low concentration (about 1/50 of the concentration of Fe oxides) and their low crystallinity, however, they are extremely difficult to identify. Therefore, most soil Mn oxides have been identified in local enrichments, such as black coatings of peds, dendrites, or nodules, which form along redox gradients. The blackish colors of Mn oxides are often separated from the brownish-orange color of Fe oxides (ferrihydrite) by extending further toward the higher redox potential, indicative of a higher mobility of Mn^{2+} as compared to Fe^{2+}.

Mn oxides are able to accumulate a wide range of other elements such as Li, Ba, As, Pb, and almost all first-row transition metals. These elements are either part of the crystal structure or they are tightly sorbed to the large surface area of Mn oxides. Mn oxides have an oxidizing potential stronger than that of O_2. Therefore, they are able to oxidize inorganic ions such as Co^{2+}, Cr^{3+}, and As^{3+}, thereby either increasing (e.g., Cr^{6+}) or decreasing (e.g., Co^{3+}, As^{5+}) their mobility and toxicity. They are also able to oxidize organic molecules, thereby enhancing the degradation of anthropogenic compounds like pesticides. Furthermore, Mn oxides catalyze condensation reactions such as the Maillard reaction. Hence, Mn oxides may play an important role in the abiotic formation of humic substances.

In spite of their importance for the geochemistry of metals and carbon, knowledge of their mineralogy and formation in soils is still relatively poor. This is due to their low crystallinity and low concentration, but also due to their complex mineral chemistry. Since microspectroscopic methods became available recently which are suited to study Mn oxides in soils, this may change within the next years.

Based on the arrangement and linkage of their metal octahedra, the Mn oxides have been categorized into three major groups: (1) phyllomanganates or layer structures, (2) tectomanganates or tunnel structures, and (3) chain structures. The most common Mn oxide minerals are listed in Table 2, and their crystal structures are shown in Figure 2.

Phyllomanganates (Layer Structures)

Phyllomanganates are commonly observed after microbially catalyzed oxidation of Mn^{2+} and they are ubiquitous in soils.

Birnessite forms in a wide variety of soils. It consists of sheets of edge-sharing MnO_6 octahedra with vacancies. The resulting negative layer charge is compensated by cations such as Na^+, K^+, Ca^{2+}, Mg^{2+}, and Mn^{2+}, which may cover the vacancies. The interlayer region contains water, resulting in a d-spacing of 7 Å (Figure 2h). Three other minerals, vernadite, chalcophanite, and buserite, are structurally very similar to birnessite. Vernadite is frequently observed as an initially amorphous oxidation product of microorganisms and has been found in soils and in

Table 2 Manganese oxide minerals

Mineral name (synthetic equivalent)	*Formula*	*Structural type*	*Figure*
Birnessite[a]	$(Na, Ca, Mn^{2+})Mn_7O_4 \cdot 2.8\ H_2O$	Layer structure	2h
Vernadite[a] (δ-MnO_2)	$MnO_2 \cdot nH_2O$	Layer structure	–
Chalcophanite	$ZnMn_3O_7 \cdot 3\ H_2O$	Layer structure	2h
Buserite	$Na_4Mn_{14}O_{27} \cdot 21\ H_2O$	Layer structure	–
Lithiophorite[a]	$LiAl_2Mn_2^{4+}Mn^{3+}\ O_6(OH)_6$	Layer structure	2e
Todorokite[a]	$(Na, Ca, K)_x(Mn^{4+}\ Mn^{3+})_6O_{12} \cdot 3.5\ H_2O\ (x = 0.3–0.5)$	3 × 3 tunnel	2i
Romanechite	$Ba_{0.66}\ (Mn^{4+}\ Mn^{3+})_5O_{10} \cdot 1.34H_2O$	3 × 2 tunnel	–
Hollandite (α-MnO_2)	$Ba_x(Mn^{4+}Mn^{3+})_8O_{16}\ (x = 1)$	2 × 2 tunnel	2f
Cryptomelane	$K_x(Mn^{4+}Mn^{3+})_8O_{16}\ (x = 1.3–1.5)$	2 × 2 tunnel	2f
Coronadite	$Pb_x(Mn^{4+}Mn^{3+})_8O_{16}\ (x = 1–1.4)$	2 × 2 tunnel	2f
Pyrolusite (β-MnO_2)	MnO_2	Rutile	2j
Manganite (γ-MnOOH)	MnOOH	Rutile	2j
Ramsdellite	MnO_2	Diaspore	2a
Groutite (α-MnOOH)	MnOOH	Diaspore	2a
Feitknechtite (β-MnOOH)	MnOOH	Boehmite	2c
Hausmannite	Mn_3O_4	Inverse spinel	2d

[a] Minerals commonly occurring in soil.

freshwater lakes. Its structure may be similar or identical to turbostratic H^+-birnessite. Chalcophanite has been observed in Zn-contaminated soils. In comparison with birnessite, it has a higher number of vacancies that are covered by Zn in octahedral or tetrahedral coordination (Figure 2h). Buserite forms in hydrothermal environments. In comparison with the former phyllosilicates, it has a larger d-spacing of 10 Å, probably due to a larger amount of interlayer water. By dehydration, buserite transforms into birnessite.

Lithiophorite has been frequently identified in soils. Its layer structure consists of sheets of edge-sharing MnO_6 octahedra which alternate with sheets of $(Al, Li)(OH)_6$ octahedra (Figure 2e). The cation sites in the Mn-O octahedral sheet are occupied with Mn^{4+} and Mn^{3+}. The net negative layer charge is counterbalanced by the positive charge of the (Al, Li)-OH octahedral sheets. Li is frequently replaced by Ni, Cu, and Zn, while Co substitutes for Mn within the octahedral Mn sheets. Asbolane forms in soils with high Co and Ni concentrations. Its structure is very similar to that of lithiophorite, with alternating sheets of Mn^{4+}-O octahedra and Co-Ni-OH octahedra (Figure 2e). The Co-Ni sheet may be discontinuous (island-like).

Tectomanganates (Tunnel Structures)

Tectomanganates are formed from double or triple chains of MnO_6 octahedra, which form tunnels (or channels). These tunnels are occupied by large foreign cations and water molecules.

Todorokite is the most frequently observed soil tectomanganate. It consists of triple chains of edge-sharing MnO_6 octahedra linked to form large 3 × 3 tunnels. These tunnels contain Na, Ca, K, Ba, Sr, and water molecules (Figure 2i). The occurrence of variable tunnel widths (3 × 2, 3 × 3, 3 × 4, and 3 × 5) observed in high-resolution transmission electron microscopy (HRTEM) images suggests that todorokite represents a family rather than a single mineral.

Hollandite, cryptomelane, and coronadite, sometimes grouped as α-MnO_2, have rarely been observed in soils. They consist of double chains of edge-sharing MnO_6 octahedra linked to form 2 × 2 tunnels (Figure 2f). The tunnels contain water molecules along with cations, which are primarily Ba^{2+} in hollandite, K^+ in cryptomelane, and Pb^{2+} in coronadite. Natural samples usually contain a variety of cations in the tunnels. The large cations are located at specific tunnel sites, and their presence is necessary to prevent the structure from collapsing. The charges of the tunnel cations are balanced by the substitution of Mn^{4+} by Mn^{3+} in the tunnel walls.

Chain Structures

Mn oxides of this group have been synthesized in the laboratory at ambient temperature and pressure, and at moderately alkaline pH. However, they have rarely been observed in soil. Pyrolusite (MnO_2) and manganite (γ-MnOOH) consist of single chains of edge-sharing MnO_6 octahedra which form 1 × 1 pseudotunnels (Figure 2j). Ramsdellite (MnO_2) and groutite (α-MnOOH) are isostructural with goethite (Figure 2a); feitknechtite (β-MnOOH) has a structure similar to lepidocrocite (Figure 2c). Hausmannite (Mn_3O_4) has a disordered-spinel structure analogous to magnetite (Figure 2d).

Al Oxides

In contrast to Fe and Mn, most Al released by the weathering of Al-containing silicates is taken up by secondary, pedogenic phyllosilicates. As long as the Si concentration in soil solution is more than 0.5 mg l^{-1}, the formation of phyllosilicates out-competes the formation of Al oxides. Furthermore, the complexation of Al^{3+} by organic matter prevents Al oxide precipitation. Therefore, Al oxides are commonly observed only in organic matter-poor soils of tropical and subtropical climate, where intense weathering causes secondary phyllosilicates to dissolve and Si is lost by leaching. The (colorless) Al oxides are either evenly distributed within the soil matrix or enriched in concretions (pisolites). Horizons cemented by Al oxides are called alcretes.

Al oxides may have a high surface area (up to 600 $m^2 g^{-1}$), a high PZC (pH 8–10), and a pH-dependent surface charge. Similar to Fe and Mn oxides, they tightly sorb heavy metals (Cu, Pb, Zn, Ni, Co, Cd) and anions (phosphate, silicate, molybdate, sulfate, catechol) by forming inner-sphere sorption complexes. The adsorption is related to the reactive surface area (singly coordinated OH^- groups at crystallite edges) rather than the total surface area. Aluminum oxides, in addition to Fe oxides, are responsible for the aggregation of Oxisols and Ultisols, but the mechanism of aggregation is unclear. Aluminum oxides have a low solubility at slightly acidic and neutral pH, but at low pH soluble species form that are toxic to plants and animals. These soluble species impact food production on naturally acidic soils and contribute to forest decline in areas influenced by acid rain.

Of six different Al oxide minerals, predominantly gibbsite and, less commonly, boehmite, nordstrandite, and bayerite form under soil conditions (Table 3). Diaspore and corundum are occasionally found in bauxite deposits, but are rare in soils. Aluminum also forms highly reactive, poorly crystalline precipitates and colloids.

Gibbsite

Of the four $Al(OH)_3$ polymorphs (Table 3), only gibbsite (γ-$Al(OH)_3$) is common in soil. It consists of sheets of edge-sharing octahedra stacked along the *c*-axis (Figure 2k). Only two-thirds of the available octahedral sites are filled with Al^{3+} ions analogous to the dioctahedral sheets of phyllosilicate clay minerals. In soil, gibbsite forms thick, platy crystals. In saprolites, feldspars (plagioclase) may transform directly into gibbsite, while maintaining the original shape of the feldspar crystals (pseudomorphs).

Table 3 Aluminum oxide minerals

Mineral	*Formula*	*Figure*
Gibbsite[a]	γ-$Al(OH)_3$	2k
Nordstrandite	$Al(OH)_3$	2k
Bayerite	$Al(OH)_3$	2k
Doyleite	$Al(OH)_3$	2k
Diaspore	α-AlOOH	2a
Boehmite	γ-AlOOH	2c
Corundum	α-Al_2O_3	2b

[a] Minerals commonly occurring in soil.

Gibbsite is found in Oxisols on old, stable upland landscapes in association with Fe oxides (goethite, hematite) and the aluminosilicate kaolinite. Intense weathering and leaching under humid tropical and subtropical climate over long periods of time (millions of years) have depleted these soils with respect to alkali metals, earth alkali metals, and Si, thereby enhancing the formation of gibbsite. Gibbsite contents of Oxisols may be as high as 600 g kg^{-1}, but some Oxisols do not contain gibbsite at all. In Ultisols and Inceptisols of tropic to temperate climates, gibbsite is derived from the underlying saprolite, which has formed by weathering of igneous and metamorphic rocks. The gibbsite content generally decreases from the saprolite to the soil surface. Finally, gibbsite has been observed in Andisols, where it formed from rapidly weathering volcanic ash. Although these soils are often very young (hundreds of years) and Si-rich, sufficient Si was leached locally to explain the formation of gibbsite.

Gibbsite formation is conditioned by the intensity of leaching which, in turn, is affected by rainfall, temperature, parent rock, topography, ground-water table, vegetation, and time. Environments with warm temperatures, high rainfall, and free drainage favor desilication and leaching of ions, as well as mineralization of organic matter. Anions such as sulfate, carbonate, phosphate, and silicate, as well as organic ligands such as citric, malic, tannic, aspartic, and fulvic acid, which have a strong affinity for Al^{3+}, may interfere with the crystallization of $Al(OH)_3$. This explains why there is no or only small amounts of gibbsite in the A horizon of acidic soils high in organic matter and exchangeable Al, while it is found in greater quantities in deeper horizons. However, gibbsite may also decrease with soil depth. A possible explanation is an inverse gradient of $Si(OH)_4^o$ (silica) activity: rain water infiltrating at the soil surface is not yet in equilibrium with silicates and therefore has a lower silica activity as compared to the soil water at a greater depth, which had more time to equilibrate. The heterogeneous distribution of gibbsite in a landscape may be similarly explained by the silica activity of percolating waters.

Rare Al Oxide Minerals

The difference between gibbsite and its other polymorphs, bayerite, nordstrandite, and doyleite, consists of a slightly different arrangement of hydroxyl groups. While gibbsite formation is favored by slow hydrolysis of Al and by a pH <6, nordstrandite and bayerite form at neutral to alkaline pH under fast hydrolysis. Correspondingly, their rare occurrences are related to limestone materials. Boehmite (γ-AlOOH), which is isostructural with lepidocrocite (Figure 2c), has been identified in lateritic materials and in bauxites. It may form from gibbsite by diagenesis or hydrothermal alteration. The oxyhydroxide diaspore (α-AlOOH) is isostructural with goethite (Figure 2a). It has been identified as a surface-weathering product formed by the desilication of kaolinitic clay. The rarely found oxide corundum (α-Al_2O_3), which is isostructural with hematite (Figure 2b), may be derived from corundum-containing parent rock or form by heating of the soil by bush fires.

Al Hydrolysis and Poorly Crystalline Al Hydroxides

Al^{3+} cations released from Al-bearing minerals are initially coordinated with six water molecules, forming the hexaquo complex $Al(H_2O)_6^{3+}$. This hexaquo complex hydrolyses in several steps, the first and most important one in soil solution being:

$$[Al(H_2O)_6]^{3+} + H_2O \Leftrightarrow [Al(OH)(H_2O)_5]^{2+} + H_3O^+$$

$$pK_s = 5.0$$

At higher Al^{3+} or OH^- concentrations, the mononuclear complexes start to polymerize to form polynuclear species. The smallest unit identified so far is $AlO_4Al_{12}(OH)_{24}(H_2O)_{12}^{7+}$. This so-called Al_{13} or Keggin polymer consists of a highly symmetrical, tetrahedrally coordinated Al center enclosed in a cage-like structure of 12 octahedrally coordinated Al centers. Further polymerization of this colloidal species leads to the precipitation of a poorly crystalline solid phase.

Si Oxides

The Si oxides differ from the oxides of Fe, Mn, and Al in several aspects. The most prevalent Si oxide, quartz (SiO_2), is a primary mineral in many rocks, contributing approx. 120 $g\,kg^{-1}$ to the average composition of the Earth's upper mantle. In soils, dissolved Si associates with Al to form aluminosilicate clay minerals. However, secondary Si oxides precipitates may also form in soil, sometimes to such an extent that they cement the soil matrix to form duricrusts and silcretes. Primary quartz crystals are very pure, hard, and recalcitrant, and occur in soil as large particles (sand). Their small surface area, their small amount of reactive bonds at the surface, and the lack of isomorphic substitution make them chemically very inert. Quartz has the lowest ion exchange capacity of all soil minerals. The solubility of Si oxides is low at usual soil pH (2–8), but strongly increases at pH >8. The main soluble species, $Si(OH)_4^o$ (monosilicic acid), is taken up by higher plants, especially grasses, to increase the mechanical strength and the resistance against fungal pathogens and herbivores, and is later recycled back into the soil as biogenic opal. Radiolarians, diatoms, silica sponges, and cyanobacteria are known to accumulate large amounts of Si oxides in fresh and marine waters, and microorganisms may exert a similar function in soils, competing with the inorganic formation of aluminosilicate clay minerals.

The Si oxides are tectosilicates. The repeating unit is a SiO_4 tetrahedron in which each O is linked to Si atoms of adjacent tetrahedra, forming a 3D framework structure. This contrasts with the Fe, Al, Mn, and Ti oxides, in which the basic unit is a cation in octahedral coordination. In the following sections, only α-quartz, α-cristobalite, and opal, which commonly occur in soil, will be described (Table 4).

Table 4 Silicon oxide minerals

Mineral	*Formula*	*Structural type*	*Occurrence*	*Figure*
α-Quartz[a]	SiO_2		Ubiquitous	2m
α-Tridymite	SiO_2			–
α-Cristobalite[a]	SiO_2		Volcanic deposits	2n
Opal-C[a]	$SiO_2 \cdot xH_2O$	Well-ordered α-crystobalite	Volcanic deposits	–
Opal-CT[a]	$SiO_2 \cdot xH_2O$	Disordered α-crystobalite and α-tridymite	Cherts, porcelanites, fossil wood, silcretes	–
Opal-A[a]	$SiO_2 \cdot xH_2O$	Nearly amorphous	Biogenic ubiquitous; inorganic in strongly weathered soils, duripans, silcretes	–

[a] Minerals commonly occurring in soil.

Quartz and Cristobalite

α-Quartz consists of paired helical chains of corner sharing SiO_4 tetrahedra that spiral along the *c*-axis. The intertwined chains produce open, hexagonal channels (Figure 2m). In cristobalite, the tunnels are larger (Figure 2n), and the density is smaller ($2.32\,g\,cm^{-3}$) than quartz ($2.65\,g\,cm^{-3}$).

Quartz is ubiquitous in soil environments, while cristobalite is restricted to soils developed from volcanic materials. The quartz in soils is mainly a primary mineral, inherited from the parent material. In soil solution, quartz may form from soluble $Si(OH)_4^o$ by polymerization, precipitation of amorphous, hydrous SiO_2, which slowly transforms into crystalline quartz. Direct precipitation from solution seems to be unlikely. Quartz is generally concentrated in the sand and silt fractions of soils, with a lower frequency in the coarse-clay fraction (0.2–2 μm). Quartz may comprise more than 90% of the inorganic fraction of Quartzipsamments. By contrast, highly weathered and leached Oxisols have only very low amounts of quartz left. The solubility of quartz ranges from 1 to $4\,mg\,l^{-1}$ Si depending on particle size.

Opal

Opal is classified into three structural groups, opal-C (well-ordered α-cristobalite), opal-CT (disordered α-cristobalite, α-tridymite) and opal-A (highly disordered, nearly amorphous). Opal contains $40–90\,g\,kg^{-1}$ water, and the specific gravity ranges from 1.5 to $2.3\,kg\,dm^{-3}$.

Opal-C and α-cristobalite occur in Andepts and other soils derived from volcanic material. Opal-CT has been identified in soils derived from bentonites, siliceous shales, and indurated silicates. Opal-CT is also a primary constituent of many cherts, porcelanites, fossil wood, silcretes, and some duripans. Opal-A may form by both organic and inorganic processes in pedogenic environments. Biogenic opal-A originates from Si accumulated by plants and aquatic organisms, and thus occurs under a wide range of environmental conditions. Biogenic opal is a nearly ubiquitous constituent of soils. Opal phytoliths are mostly derived from grasses, while sponge spicules, diatoms, and radiolaria, which form in aquatic environments, may reach soils via weathering of sediment bedrocks. Amounts of opal phytoliths in soils commonly range from less than 1 to $30\,g\,kg^{-1}$, usually occurring in the 5- to 50-μm size fraction, and decreasing with soil depth. Inorganic opal-A forms from supersaturated soil solution. It occurs as nodules and as the primary cement of indurated soil horizons (duripans) and silcretes.

At ambient temperature and neutral pH, the solubility of amorphous Si is approximately 50–60 mg $Si\,l^{-1}$, and therefore much higher than that of quartz. The presence of organic molecules greatly enhances the dissolution rates of Si oxides. Coatings of chemisorbed Al and Fe, however, reduce the dissolution rates. Oxides of Fe and Al, by acting as a sink for soluble Si, greatly increase the dissolution rates of amorphous Si oxides and of reactive, uncoated quartz surfaces. Dissolution (weathering) of Si oxides, which starts below approximately $3\,mg\,Si\,l^{-1}$, is induced by the reduction of Si levels in soil solution by leaching and plant uptake.

Ti Oxides

Titanium oxides are common in many igneous and metamorphic rocks and sediments. Due to their resistance against weathering, they are often inherited to soils. However, Ti oxides also form in soil after weathering of less-resistant, Ti-bearing minerals (biotite, amphibole). Titanium oxide minerals are relatively heavy (density more than $2.9\,g\,cm^{-3}$). Their concentration in soils is usually low; because of their small surface area, they do not contribute significantly to soil reactivity. Only in tropical soils, where they become increasingly enriched during pedogenesis, has some influence on phosphate and arsenate retention been observed. Old sand dunes may contain enough Ti oxides to be mined as a commercial source for TiO_2, which is used to produce white paint. Due to their recalcitrant nature, Ti oxides can be used as reference minerals to study weathering and soil genesis.

Titanium occurs primarily in octahedral coordination, and like the Fe and Mn oxides, the structures of the various Ti oxide minerals can be described in terms of the different arrangement of the Ti-containing octahedra. The principal Ti minerals are listed in Table 5. Rutile (TiO_2) is isostructural with pyrolusite and manganite, consisting of single chains of edge-sharing TiO_6 octahedra (Figure 2j). Rutile is a common residual mineral occurring in the sand and silt fractions of a variety of soils.

Table 5 Titanium oxide minerals

Mineral	*Formula*	*Figure*
Rutile[a]	TiO_2	2j
Anatase[a]	TiO_2	2l
Ilmenite[a]	$Fe^{2+}\,TiO_3$	2b
Pseudorutile[a]	$Fe_2O_3 \cdot nTiO_2 \cdot mH_2O$ ($n=3–5$, $m=1–2$)	–
Ulvöspinel	Fe_2TiO_4	2d
Titanomagnetite	$Fe_{3-x}\,Ti_xO_4$	2d
Titanomaghemite	$Fe_{2-x}\,Ti_xO_3$	2d

[a] Minerals commonly occurring in soil.

Anatase consists of edge- and corner-sharing octahedra that outline a three-dimensional framework, rather than distinct chains (Figure 2l). The platy crystals found in soils are believed to be a weathering product of titanite ($CaTiSiO_5$) and Ti-bearing silicates. Anatase is easily synthesized at room temperature. Ilmenite ($FeTiO_3$) is almost isostructural with hematite, with one-half of the octahedral centers occupied by Fe^{2+}, the other half by Ti^{4+} (Figure 2b). By oxidation of structural Fe^{2+} to Fe^{3+}, ilmenite weathers to pseudorutile ($Fe_2O_3 \cdot n\,TiO_2 \cdot mH_2O$; $3 < n < 5$ and $1 < m < 2$), a structurally disordered and poorly characterized mineral. Pseudorutile has been widely identified in soils, primarily as the weathering product of ilmenite. Ilmenite itself is relatively unstable in soils, because it easily weathers to pseudorutile and mixtures of rutile, anatase, and Fe oxides.

Titanomagnetites ($Fe_{3-x}\,Ti_xO_4$) are solid solutions of magnetite with ulvöspinel (Fe_2TiO_4), with Ti^{4+} occupying only octahedral sites of the magnetite structure (Figure 2d). Their formation is probably related to that of magnetite. Titanomaghemites ($Fe_{2-x}\,Ti_xO_3$) form by weathering of titanomagnetites. Their composition is between those of magnetite, maghemite and ulvöspinel. Many aspects of the Ti oxides, such as whether the minerals are relict or pedogenic, and the conditions responsible for their weathering and formation in soils, are still unresolved.

Further Reading

Banfield JF and Nealson KH (eds) (1997) Geomicrobiology: interactions between microbes and minerals. In: *Reviews in Mineralogy*, vol. 35. Washington, DC: Mineralogical Society of America.

Bigham JM, Fitzpatrick RW, and Schulze DG (2002) Iron oxides. In: Dixon JB and Schulze DG (eds) *Soil Mineralogy with Environmental Application*pp. 323–366. Madison, WI: Soil Science Society of America.

Cornell RM and Schwertmann U (2003) *The Iron Oxides: Structure, Properties, Reactions, Occurrence and Uses*, 2nd edn. Weinheim, Germany: VCH Verlagsgesellschaft.

Dixon JB and Schulze DB (2002) *Soil Mineralogy with Environmental Applications*. Madison, WI: Soil Science Society of America.

Dixon JB and Weed SB (1989) *Minerals in Soil Environments*. Madison, WI: Soil Science Society of America.

Heaney PJ, Prewitt CT, and Gibbs GV (1994) Silica: physical behavior, geochemistry and materials applications. In: Haney PJ (ed.) *Reviews in Mineralogy*, vol. 29, pp. 375–443. Washington, DC: Mineralogical Society of America.

Kämpf N, Scheinost AC, and Schulze DG (2000) Oxide minerals in soils. In: Sumner ME (ed.) *Handbook of Soil Science*pp. 125–168. Boca Raton, FL: CRC Press.

Waychunas GA (1991) Crystal chemistry of oxides and oxyhydroxides. In: Lindsley DH (ed.) *Reviews in Mineralogy*, vol. 25, pp. 11–68. Blacksburg, VA: Mineralogical Society of America.

METALS AND METALLOIDS, TRANSFORMATION BY MICROORGANISMS

S M Glasauer and T J Beveridge, University of Guelph, Guelph, ON, Canada
E P Burford, F A Harper, and G M Gadd, University of Dundee, Dundee, UK

Introduction

Fungi and bacteria from all major taxonomic groups have been found in metal-rich habitats such as mineral ores, acidic soils, or polluted environments. The survival of these microorganisms may depend on a number of biochemical, structural, and physiological properties that modify metal bioavailability and toxicity, which are constitutive to a species and/or have been selected by adaptation.

All microorganisms have essential metabolic and structural requirements for specific metals in metabolism, including K, Na, Mg, Ca, Mn, Fe, Cu, Zn, Co, and Ni, while some metals, e.g., Rb, Cs, Al, Cd, Ag, Au, Hg, and Pb, and metalloids, e.g., As and Te, have no apparent essential function. Intense competition for some metals, coupled with low bioavailability, has driven the evolution of tightly regulated mechanisms for metal acquisition. These include highly specialized, metal-specific membrane transport systems and metal chelates to scavenge metals effectively from the environment. All metals and metalloids, regardless of their metabolic function, are, however, toxic in excess and can inhibit growth, affect sporulation and adhesion, induce morphological changes, and affect biochemical activities. This potential toxicity has led cells to develop diverse mechanisms of defense. Both genetically coded mechanisms of resistance and nonspecific mechanisms such as oxidation or reduction of the metals to

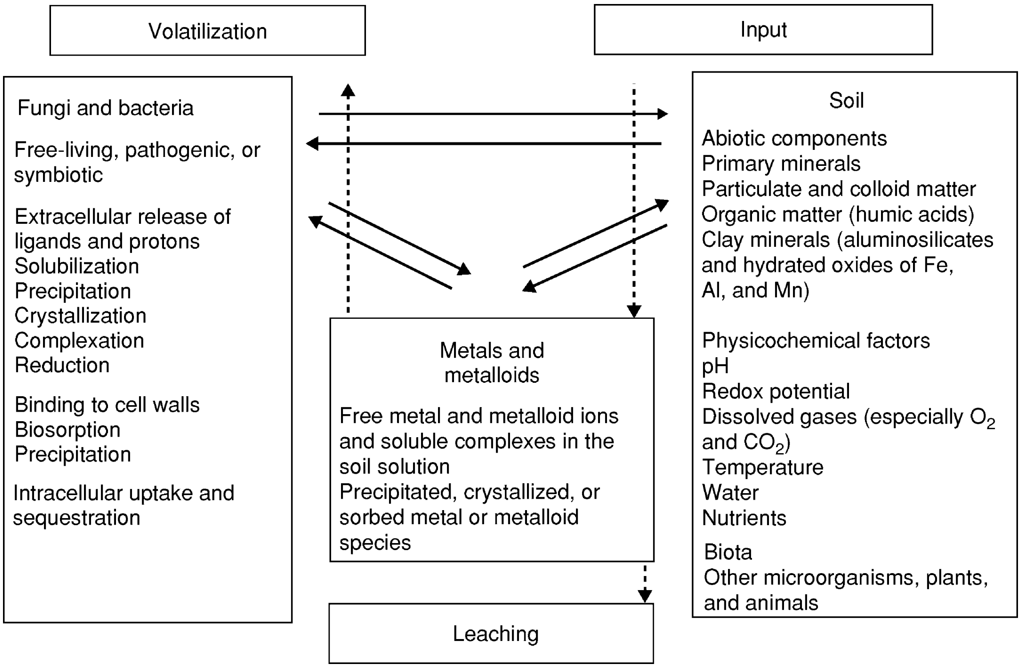


Figure 1 Fungal interactions with metals and metalloids in the soil environment. Metal(loid) species lost from the soil environment are indicated by broken arrows. Interactions mediated by microorganisms are indicated by solid arrows.

produce less toxic chemical species, or sorption and precipitation on cell surfaces, protect cells from toxic metals.

The chemical form, or species, of metals is also affected by microbial processes other than those related to nutrition and toxicity. Metal speciation is important because different metal species can have widely differing chemical and biological properties, including mobility in the environment and toxicity. Excreted cellular metabolites transform metal compounds by nonspecific reactions such as dissolution by protons and organic acids. Some bacteria use metals as electron donors and acceptors during respiration, a process that is a major pathway for the cycling of metals in soils and sediments. Such microbial transformations both mobilize and precipitate metals, leading to alteration and modification of soils and sediments. Many of these processes also have potential for the bioremediation of metal-contaminated sites.

Microbially Induced Metal Transformations

Interactions between microorganisms with metals and soil components take many forms; unravelling their complexity is integral to understanding metal speciation in the soil. The relative role of fungi and bacteria depends on associative biota (other microorganisms, plants, and animals) as well as on abiotic factors such as physicochemical components of the environmental matrix, e.g., pH, water, inorganic and organic ions, molecules, compounds, colloids, and particulates (Figure 1).

The ability of microorganisms to affect metal speciation in the soil environment stems from two opposing phenomena: mobilization and immobilization, which influence the distribution of metal species between soluble and insoluble phases. The solubilization of metals can be achieved by various mechanisms: protonation, chelation, and chemical transformation. Metal immobilization can occur by precipitation or crystallization of insoluble organic or inorganic compounds, or by sorption, uptake, and intracellular sequestration. In addition, redox reactions can mobilize or immobilize metals, depending on the metal species.

Mobilization

Fungi and bacteria can facilitate the transformation of metals into soluble forms via the dissolution of metal compounds, including oxides, phosphates, sulfides, and more complex mineral ores, or by desorption from exchange sites on clay minerals or organic matter. Metal leaching by Fe- and S-oxidizing bacteria is discussed further on. A number of other mechanisms by which microbes can leach metal ions have been proposed.

Acidification

Competition between protons and the metal (in a metal–anion complex) leads to protonation of the anion under conditions of increased acidity, and results in the release of free metal cations. Fungi and bacteria can acidify their environment in a number of ways; for example, proton efflux via plasma membrane H^+-ATPases, maintenance of charge balance, or as a result of respiratory carbon dioxide accumulation.

Complexation by Ligands

Microorganisms produce a number of extracellular metabolites that can complex metals in solution, including polysaccharides, pigments, organic acids, and siderophores. Organic acids are important in energy production as intermediates in the tricarboxylic (TCA) cycle and in cellular pH homeostasis. They can correct internal perturbations in charge balance and pH resulting from nutrient uptake by their ability to produce and consume protons via the formation and removal of carboxylic groups, and by efflux from the cell.

Citric and oxalic acids are the most commonly reported organic acids in soils and are released into the soil by fungal hyphae, lichens (a symbiotic relationship between fungi and a photosynthetic cyanobacterium or alga), bacteria, and plant rootlets (mycorrhizae). Organic acids are believed to have an essential role in nutrient acquisition as a result of their acidification and metal-binding properties. For example, phosphate is released from common soil compounds such as $CaHPO_4$, $Ca_3(PO_4)_2$, $FePO_4.4H_2O$, and $AlPO_4$, either by direct exchange of chemically binding species, or ligands, between phosphate and the organic anion or by binding of the metal to the anion. Metal complexation depends on metal-to-anion availability, the stability of the particular metal–ligand complex, and pH. Anion concentration (which will depend on the pK_a of the acid) is important, because fully protonated organic acids chelate metals poorly. The affinity depends on the stability constant of the organic acid–metal complex. For example, the efficiency of organic acids to solubilize phosphate from soils often follows the order: citrate > oxalate > malate > acetate. The bioavailability of a solubilized metal–ligand complex will depend ultimately on its long-term fate in the environment, including sorption on to components of the soil matrix or potential utilization by biota. For example, metal citrates can be highly resistant to microbial degradation, which can result in effective leaching of metals from soil. In contrast, metal oxalates crystallize.

Many soil bacteria, fungi, and lichens are particularly efficient at solubilizing metal phosphates and other metal-bearing compounds, which has been attributed to their ability to produce protons and organic ligands. This can be tested by simple laboratory experiments: if insoluble metal compounds are incorporated into solid growth medium, clear zones may appear around microbial colonies if they are capable of mineral dissolution.

In addition, organic acid production by microbes is an important agent of mineral deterioration, playing a role in both biogenic chemical weathering and soil formation. Fungi can also physically weather minerals by etching and fracturing mineral grains during hyphal growth.

Siderophores

Siderophores are the largest class of known compounds that can bind and transport, or shuttle, Fe. They are highly specific Fe(III) ligands (formation constants often greater than 10^{30}). These low-molecular-weight coordination molecules are excreted by a wide variety of fungi and bacteria to aid Fe assimilation. Although the mechanism could be used to acquire other metals, Fe is the only known essential element for which these specific organic shuttles operate. This is most likely because Fe is needed in larger amounts by cells than other poorly soluble metals, and, given the low solubility-product constant of ferric hydroxide (less than 10^{-38}), the concentration of free Fe^{3+} is too low to support microbial growth at pH values where most life exists.

In a typical soil, many different species, including plants, compete for Fe via siderophores. Organisms have most likely evolved mechanisms to ensure that Fe demand is met through the production of species-specific siderophores, or attachment to solid mineral Fe (e.g., Fe oxides) to shorten the pathway between the Fe substrate and the site of uptake.

Biomethylation

Methylation of Hg, As, Se, Sn, Te, and Pb is mediated by a range of microorganisms, including clostridia, methanogens, and sulfate-reducing bacteria under anaerobic conditions, and principally fungi under aerobic conditions, such as *Penicillium, Cephalosporium*, and *Alternaria* spp. as well as a variety of bacteria, including pseudomonads. Methyl groups are enzymatically transferred to the metal, and a given species may transform a number of different metal (loid)s. Methylated metal compounds formed by these processes differ in their solubility, volatility, and toxicity. Volatile methylated species, e.g., $(CH_3)_2Se$ and $(CH_3)_2Se_2$, are often lost from the soil.

Redox Reactions

Microorganisms mobilize metals, metalloids, and organometallic compounds by reduction and oxidation processes, exemplified by pathways for metal

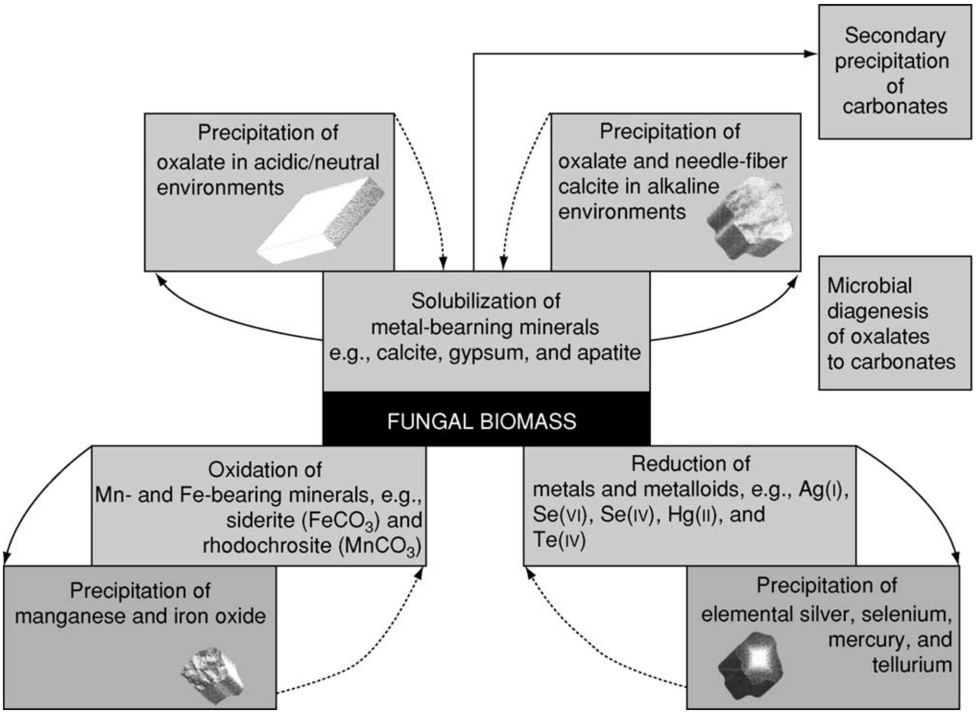


Figure 2 Metal precipitation by microorganisms. Solid arrows represent precipitation by fungi and bacteria. Broken arrows represent precipitation and association with microbial biomass (living or dead).

cycling by fungi (Figure 2). For example, oxidation of metal-complexing dimethylsulfide, dimethylsulfoxide, or thiosulfate increases metal availability if metal sulfates are formed, and the solubilities of Fe and Mn increase upon reduction of Fe(III) to Fe(II) and Mn(IV) to Mn(II). While many microorganisms reduce small amounts of Fe during metabolism, most Fe reduction is carried out by specialized anaerobic bacteria such as *Shewanella* spp. and *Geobacter* spp. that use Fe for respiration, and there is evidence that reduction of Fe(III) was one of the first forms of microbial respiration. In this process, oxidized forms of metals are used as the final electron acceptor in place of the oxygen consumed during aerobic respiration. The dissimilatory metal-reducing bacteria respire and thereby reduce oxidized metal species at or near the outer cell wall. The membrane-bound electron transport system provides energy to cells by cycling electrons through an electrochemical cascade of metalloenzymes, expelling e^- and H^+ from the cell. Energy is generated by the flux of H^+ into the cell through specific membrane-bound channels (Figure 3). Dissimilatory metal-reducing bacteria can use a variety of metal(loid)s with an appropriate redox couple, including Fe(III), Mn(IV), Se(IV), Cr(VI), and U(VI). While Fe and Mn increase their solubility upon reduction, the solubility of other metals such as U(VI) to U(IV) and Cr(VI) to Cr(III) decreases, resulting in immobilization (see Precipitation of Metals and Minerals, below). Reduction of Hg(II) to Hg(0) by bacteria and fungi results in diffusion of elemental Hg out of cells, and conferment of Hg resistance on organisms that possess this property since Hg(0) is considerably less toxic than Hg(II).

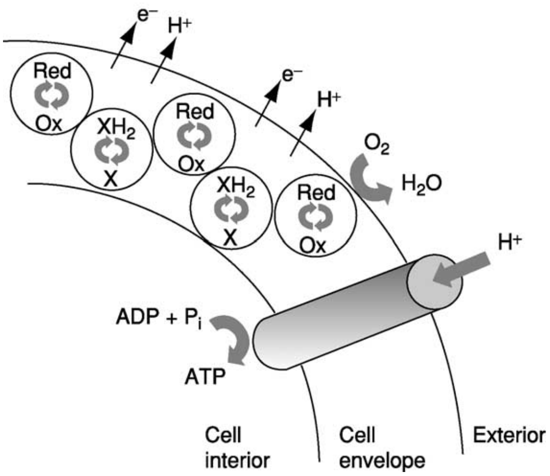


Figure 3 Simplified bacterial electron transport chain, showing linked carriers in the cell wall that cycle electrons down an electrochemical gradient to the final electron acceptor (in this case O_2). Energy is generated by the flux of protons through specific channels across the membrane. Red, reduction; Ox, oxidation; P_i, inorganic phosphorus; ADP, adenosine diphosphate; ATP, adenosine triphosphate.

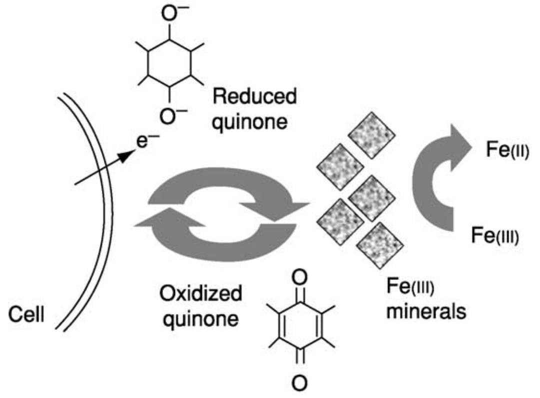


Figure 4 Proposed mechanism for electron transfer from metal-reducing cells to oxidized Fe minerals via chemical reduction of a quinone shuttle. The shuttle regenerates when Fe(III) is reduced to Fe(II).

Dissimilatory metal reduction requires that bacteria access relatively large amounts of insoluble metals. Bacteria can use redox-active organic compounds to shuttle electrons between cells and the metal being respired under anaerobic conditions. Many quinoid compounds, common components of humic substances, can function as electron shuttles. These compounds move electrons from the cell membrane to the target metal, such as iron contained in ferric oxides. The shuttle molecule regenerates when the metal is reduced and the cycle continues (Figure 4). While the use of electron shuttle mechanisms appears to be widespread among bacteria, the impact and ecologic significance of the process in natural subsurface environments are not fully understood.

In the absence of processes that mix water columns, sediments, and soils, redox potentials tend to decrease with depth, reflecting a change in the dominant electron acceptors from O_2 to NO_3^-, Mn, Fe, and (finally) S. The interplay between microbial and abiotic metal reduction forms characteristic element profiles (Figure 5). *Shewanella*, *Geobacter*, and *Desulfovibrio* spp. have been intensively studied for their metal-reducing capability. Molecular studies reveal that many as yet uncultured species exist in the subsurface which can also contribute, and improved culturing methods are likely to show novel dissimilatory metal reducers amongst these organisms. The poorly crystalline iron fraction is generally considered to be the 'reactive' fraction for these bacteria; however, iron in some clay minerals and crystalline Fe oxides could be available for reduction.

Immobilization

There are a number of processes whereby microbial biomass can immobilize metals in the soil environment. Although immobilization reduces the external free metal activity, it may also shift the equilibrium to release more metal into the soil solution.

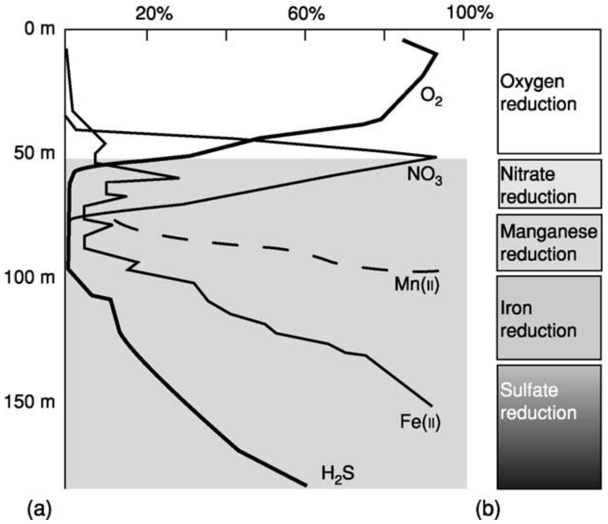


Figure 5 Reductive processes in the environment: (a) representative electron acceptor profiles (percentage of maximum) with water depth in meters. Data are from the Black Sea; (b) major processes occurring in the redox zone with depth. (Reproduced with permission from Nealson KH and Saffarini D (1994) Iron and manganese in anaerobic respiration: environmental significance, physiology and regulation. *Annual Review of Microbiology* 48: 311–43. © Annual Reviews, Inc.)

Biosorption

The strong interaction of metals with cell surfaces stems from cell wall characteristics, including component pigments and polysaccharides. Bacterial cells depend on the diffusion of nutrients for survival, and their small size, shape, and characteristic membrane structures have evolved to maximize their interaction with nutrient solutes. There is remarkable similarity in the general format of bacterial wall structure, which falls into two broad classes based on reaction with Gram's stain, which utilizes crystal violet. Gram-positive walls consist of a thick matrix of peptidoglycan and secondary polymers, which provides structural support and gives the cell its shape and form (Figure 6). The phosphate and carboxyl groups within the wall fabric and exposed at the surface confer pH-dependent negative charge to the cell in the types of environments where microbes generally exist. The point of zero charge (PZC), or pH where positive charge equals negative charge, is approximately pH 2–3 for bacterial cell walls. This charge provides the basis for metal sorption. At pH values above the PZC, the net charge on the cell wall is negative, which facilitates sorption of positively charged metal species. Gram-negative cells are also negatively charged, although structurally more

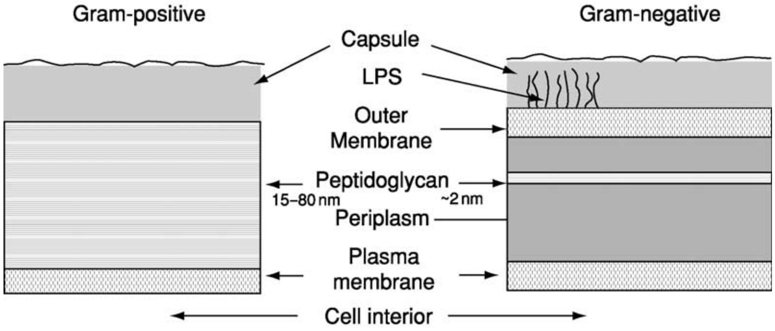


Figure 6 Cross-section view showing the cell wall structures characteristic for Gram-positive (left) and Gram-negative (right) bacteria. LPS, lipopolysaccharide.

complex. In this case, a thin layer of peptidoglycan floats in a gel-like periplasm, sandwiched between the plasma membrane and the outer membrane (Figure 6). The outer membrane has characteristic surface lipopolysaccharides (LPS) containing ionizable phosphoryl, carboxyl, hydroxyl, and amine groups; the phosphoryl groups are the most reactive to metals. In addition to these basic structures, surface capsules, sheaths, and S-layers provide additional sites for metal interaction and the heterogeneous nucleation of minerals.

Fungal cell walls are predominantly composed of polysaccharides (80–90%) with glycoproteins and some lipids. The polysaccharide component includes microcrystalline fibrils of β-linked polysaccharides (chitin, chitosan, and β-glucans) that are responsible for wall strength and rigidity. As for bacteria, the chemical properties of the functional groups associated with fungal walls, including carboxyl, amine, hydroxyl, phosphoryl, and sulfhydryl groups, provide the basis for the attraction of metals to cell walls. In addition, pigments such as melanin contribute to rigidity and have significant metal-binding properties, preventing metal entry into melanized structures.

Precipitation of Metals and Minerals

Partitioning in the cellular environment Metals are found in all parts of bacterial and fungal cells: in the cytoplasm, the cell wall, or sorbed to the cell surface. At sufficient concentrations, metal precipitates begin to develop or nucleate on or within cells, or in the immediate extracellular environment (demonstrated for a bacterial cell in Figure 7). In addition to immobilizing metals, this also reduces the external free metal activity and may shift the equilibrium to release more soluble metal to the soil. Precipitation and crystallization restrict metal leaching and bioavailability, and can potentially release nutrients such as sulfate and phosphate.

The formation of extracellular and sorbed precipitates depends on the chemical composition of the extracellular environment and is often mediated by cellular processes. The precipitates can be copious and may completely enmesh the cells, effectively mineralizing them (Figure 8). Environmental anions, such as OH^-, SO_4^{2-}, PO_4^{3-}, and CO_3^{2-}, provide counterions for mineral hydroxides, sulfate and sulfides, phosphates, carbonates, and oxalates. It is possible that cellular metabolism and excretion of such anions also contributes to these minerals. Metal oxide and hydroxide minerals formed by abiotic reactions also attach to cells, and it can be difficult to distinguish these phenomena (Figure 9). This is, however, another important mechanism by which cells can concentrate metals.

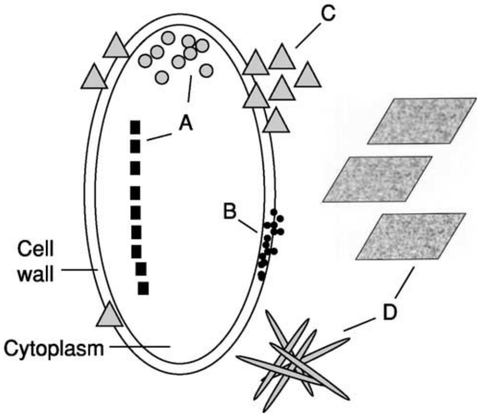


Figure 7 Possible locations of metal precipitates and minerals associated with a bacterial cell: A, biogenic precipitates within the cell cytoplasm; B, metal precipitates formed by nucleation in the cell wall, e.g., periplasmic; C, mineral nanocrystals formed independently of cell activities and subsequently sorbed to the cell wall; D, minerals formed outside cells due to the influence of cellular activities.

Inside cells, metals can be built into cellular constituents. Excess or toxic metals may be effluxed, form intracellular precipitates, bind to metal-complexing molecules, and, in eukaryotes, partition into vacuoles. The formation of internal precipitates often results from a change in metal oxidation state, which does not require an active cell; for example, gold is

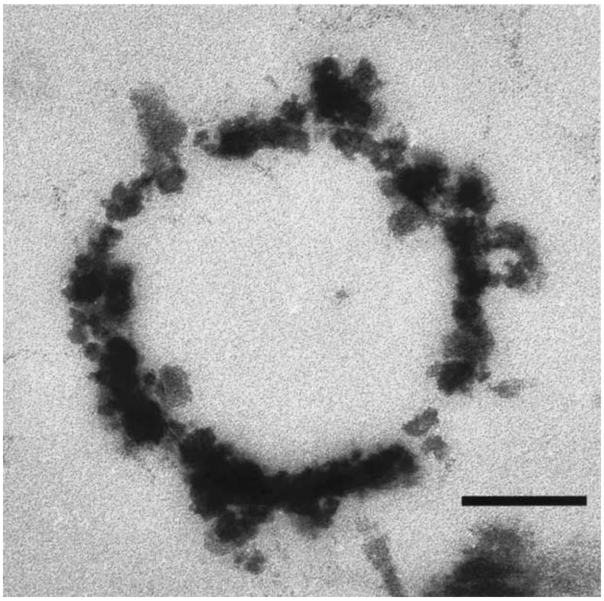


Figure 8 Transmission electron micrograph of a thin-section sample taken from a Yellowstone hot spring, showing an unknown bacterial cell embedded in metal-rich precipitates. Scale bar 150 nm.

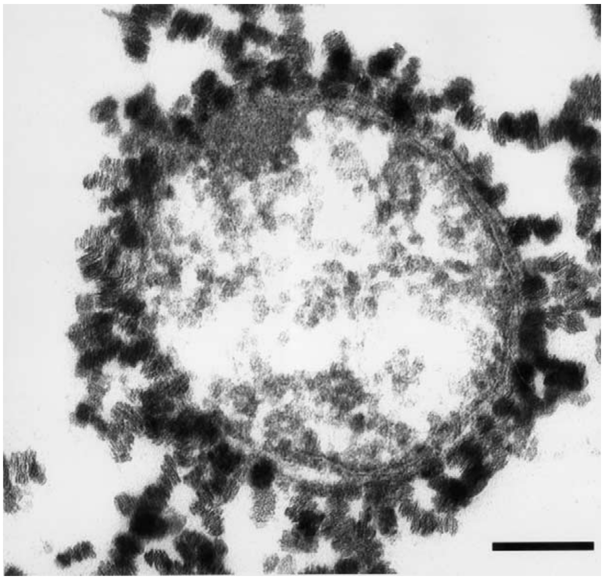


Figure 9 Transmission electron micrograph of a thin section showing abiotically precipitated goethite nanocrystals sorbed to the cell wall of *Shewanella putrefaciens*. Scale bar 150 nm.

highly toxic to cells, yet intracellular precipitates of elemental gold form in bacteria after cells are killed. The formation of oriented chains of magnetite (Fe_3O_4) crystals by magnetotactic bacteria, found in sediments and hydric soils, is a unique example of directed, intracellular mineral precipitation. In this case, the formation of mineral precipitates is a mechanism that enables the bacteria to orient in the Earth's magnetic field.

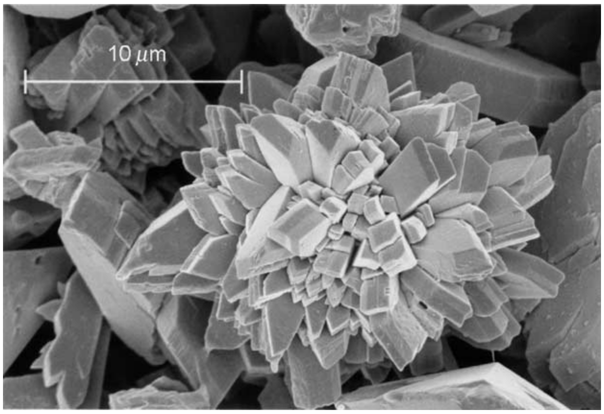


Figure 10 Calcium oxalate precipitated by *Aspergillus niger*. Scale bar 10 μm.

Oxalates Calcium oxalate is the most common form of oxalate associated with soils and leaf litter, occurring as dihydrate (weddellite) or the more stable monohydrate (whewellite). Calcium oxalate crystals are commonly associated with free-living, pathogenic, and plant-symbiotic fungi and are formed by the reprecipitation of solubilized calcium as the oxalate. Fungal-derived calcium oxalate can exhibit a variety of crystalline forms (tetragonal, bipyramidal, plate-like, rhombohedral, or needles), which often associate with the outer surfaces of hyphae (Figure 10). The formation of calcium oxalate by fungi has an important influence on biological and geochemical processes in soils, acting as a reservoir for calcium in the ecosystem, and also influencing phosphate availability.

Fungi can also produce other metal oxalates with a variety of different metals and metal-bearing minerals, e.g., Cd, Co, Cu, Mn, Sr, and Zn. The formation of metal oxalates may provide a mechanism whereby fungi can tolerate environments containing potentially high concentrations of toxic metals. A similar mechanism occurs in lichens growing on copper sulfide-bearing rocks, where precipitation of copper oxalate occurs within the thallus.

Carbonates In many arid and semiarid regions, calcareous soils and near-surface limestones (calcretes) are often secondarily cemented with calcite ($CaCO_3$). This phenomenon has been partly attributed to physicochemical processes; however, the abundance of calcified fungal filaments in weathered profiles of chalky limestone and Quaternary calcretes indicates fungal activity. Calcite precipitation on fungal hyphae has also been observed with the ectomycorrhizal fungus *Pisolithus tinctorius* in calcareous soil.

Calcite has been proposed to form indirectly through fungal excretion of oxalic acid and precipitation of calcium oxalate, which results in dissolution of the internal pore walls of the limestone matrix. This enriches the matrix solution in free carbonate ions. During passage of the solution through the pore walls, calcium carbonate recrystallizes as a result of a decrease in CO_2, hardening the limestone. Calcite crystals act as sites of further secondary calcite precipitation when the hyphae decompose. Through these processes, fungal activity is important in lithification (hardening), alteration, and diagenesis of subsurface limestone formations.

Mineralized carbonate precipitates are also found in association with bacterial biofilms. One striking example of microbial carbonate precipitation is the towering microbialites at Lake Van (Turkey), formed by thick biofilms of cyanobacteria (*Pleurocapsa* spp.) that colonize crusts of inorganic calcite precipitated in the alkaline soda lake water. Groundwater forced up and through the carbonate matrix raises the towers of aragonite and calcite to heights up to 40 m.

Redox Reactions

Bacterial and fungal oxidation The range of bacterial species with the capability to oxidize Fe and Mn spans sedimentary and soil environments. For bacteria to use Fe(II) specifically for growth, however, they must recover the energy produced in the oxidative process. Demonstration of a respiration-linked, energy-yielding reaction is particularly difficult for iron due to its rapid chemical oxidation at neutral pH and atmospheric O_2. It has been more readily shown for bacteria that grow at low pH and/or low O_2, conditions that drastically slow chemical oxidation of Fe(II). Bacterial Fe oxidation is probably ubiquitous in environments with sufficient Fe^{2+} and conditions to support the growth of bacteria that can make a living by oxidative respiration of reduced metal species. Specific environments include drainage waters and tailing piles in mined areas, pyritic and hydric soils (bogs and sediments), drain pipes and irrigation ditches, and plant rhizosphere zones.

Iron-oxidizers found in acidic soil environments are acidophilic chemolithotrophs, organisms that use inorganic sources of chemical energy, such as *Thiobacillus ferrooxidans*, which is significant for its role in generating acid mine drainage. Facultative chemolithotrophic microaerophiles such as *Leptothrix ochracea*, *Sphaerotilus natans*, and *Gallionella ferruginea* are common in mildly acidic to neutral environments, and accumulate abundant Fe hydroxide precipitates on their filament or stalk structures that can age to more crystalline minerals.

After Fe, Mn is the second most abundant transition metal in the Earth's crust and is actively cycled in the same environments as Fe. Mn oxidation is carried out by many of the organisms that are implicated in Fe oxidation, including *Bacillus*, *Pseudomonas*, and *Leptothrix* spp. In contrast to Fe^{2+}, substantial chemical oxidation of Mn^{2+} does not occur below pH 8, and solutions containing Mn^{2+} can be stable for weeks. This fact has allowed bacterial Mn oxidation to be identified more readily than its Fe counterpart.

It is not yet clear why cells carry out metal oxidation if not for gaining energy. Oxidation may help the cell rid itself of toxic intracellular concentrations of the metal, or the precipitates that coat the bacterial surface subsequent to oxidation may protect the cells from attack by other microbes. Protection from UV radiation would have been especially crucial during the early evolution of life on Earth.

Fungi also oxidize metals in their environment. Desert varnish is an oxidized metal layer (patina) a few millimeters thick found on rocks and in soils of arid and semiarid regions, and is believed to be of fungal and bacterial origin. *Lichenothelia* (a fungus) can oxidize manganese and iron from metal-bearing minerals such as siderite ($FeCO_3$) and rhodochrosite ($MnCO_3$), and metals from rainfall or windblown dust.

Dissimilatory metal-reduction by bacteria When dissimilatory metal-reducing bacteria utilize relatively soluble metals such as Cr(VI) and U(VI) as electron acceptors, less-soluble Cr(III) and U(IV) species may result. Once reduced, soluble metals sorb to the cells or partition to the bulk extracellular solution, where new minerals can form from the metals and suitable counterions. Postreduction partitioning of metals depends on the element. For example, U(IV) accumulates as fine precipitates in the periplasm of Gram-negative metal reducers (Figure 11). Fe^{2+} and Mn^{2+}, generated by dissimilatory reduction, can form copious extracellular precipitates and nodules. The fate of the reduced metal depends on the environment, and it is feasible that a metal such as Fe, along with any associated redox-active metals, could be cycled hundreds of times between reducing and oxidizing zones before being incorporated into refractory minerals (Figure 12).

Fungi and other microorganisms also precipitate reduced forms of metal species, e.g.: Ag(I) to Ag(0); selenate (Se(VI)) and selenite (Se(IV)) to Se(0); and tellurite (Te(IV)) to Te(0), by a variety of mechanisms.

Biotechnological Applications of Microbial Metal Transformations

Ore Leaching

Metals in ores that are not worth recovering by smelting can sometimes be effectively leached by

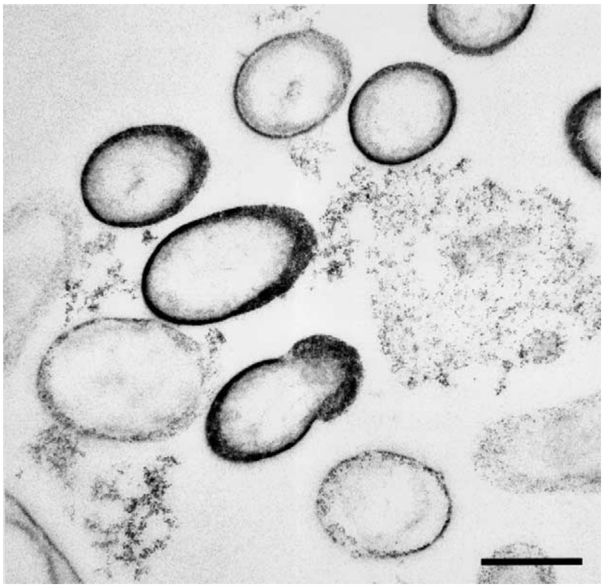


Figure 11 Transmission electron micrograph of a thin section showing uranium precipitates formed in the periplasmic space of *Geobacter metallireducens* cells during dissimilatory reduction of U(VI). No metal stains were used, so that the contrast in the image is from the uranium precipitates only. Scale bar 150 nm.

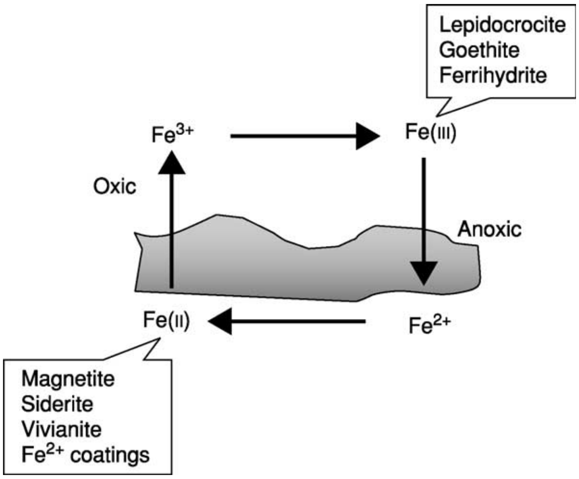


Figure 12 Cycling of Fe between soluble and insoluble mineral, oxidized and reduced forms mediated by microorganisms in soils and sediments. Although by quantity Fe is the major element to undergo redox transformations, other metals can be cycled by the same processes and associate with the precipitating minerals.

bacteria and fungi. Important species of bacteria in such processes are Fe and S oxidizers such as *Thiobacillus ferrooxidans*, *T. thiooxidans*, and *Leptospirillum ferrooxidans*. Their activities solubilize Fe and S contained in metal sulfides, with the concomitant release of associated metals such as Cu and Zn. Large amounts of Fe^{2+} and sulfuric acid are generated in the process. It has been accepted most commonly that bacteria catalyze the oxidation of Fe^{2+} to regenerate Fe^{3+}, but do not directly attack the mineral:

$$14Fe^{2+} + 3.5O_2 + 14H^+ \rightarrow 14Fe^{3+} + 7H_2O \quad \text{bacteria} \qquad [1]$$

$$FeS_2 + 8H_2O + 14Fe^{3+} \rightarrow 15Fe^{2+} + 2SO_4^{2-} + 16H^+ \quad \text{abiotic} \qquad [2]$$

where eqn [1] gives the reaction for bacteria, and eqn [2] is the abiotic reaction. There is, however, increasing evidence that *T. ferrooxidans* may directly enhance dissolution of metal sulfides via an enzymatic mechanism. Approximately 90 mol Fe^{2+} is oxidized to assimilate 1 mol C, which explains the tremendous amount of acidity – 'acid mine drainage' – that is generated. Industry can take advantage of the acidic leachate by recirculating it through the ore body to dissolve and harvest more metal. Here, the acidic leachate is recirculated through the ore body to solubilize more metal, and the soluble metals are then separated by sedimentation, solvents, or electrodialysis. When uncontrolled, such as at poorly managed or abandoned mining sites, the same process can have disastrous environmental consequences.

Extracellular ligands produced by fungi, including *Aspergillus* and *Penicillium* spp., have also been used to leach metals such as Zn, Cu, Ni, and Co from a variety of solid materials, including low-grade mineral ore, solid waste, and electronic scrap materials.

Bioremediation

Toxic metals in natural, industrial, and agricultural soils may be a risk to biota at all trophic levels if they exist at potentially toxic levels and are, or become, bioavailable. Some of the processes outlined above have potential for application to treat contaminated land. Solubilization processes provide a route for removal of metals from soil matrices, whereas immobilization processes enable metals to be transformed to insoluble chemically inert forms. Although some processes could be used *in situ* (e.g., leaching using S-oxidizing bacteria), many are probably most suitable for *ex situ* use in bioreactors, where the mobilized or immobilized metal can be separated from soil components. Living or dead fungal and bacterial biomass and metabolites have been used to remove metals and metalloids from solution by biosorption or chelation.

Recently there has been renewed interest in the ability of some mycorrhizal fungal symbionts to improve plant metal tolerance and increase metal translocation to the aboveground parts of plants. Such associations could be used to enhance the phytoextraction capabilities of plants grown for removing toxic metals from soil by accumulation in their shoots. Amelioration of metal phytotoxicity by mycorrhizal fungi has been widely demonstrated, particularly when inoculated with mycorrhizae isolated from metalliferous or acid soils.

The increased attention to microbial activities in anaerobic, subsurface environments in the past few years has opened up new possibilities for the bioremediation of metal contaminants. Metal(loid)s that form insoluble precipitates when reduced may be of particular interest for *in situ* treatment, such as Se(o), Cr(III), Tc(IV), and U(IV). Natural processes in subsurface soil may immobilize contaminants otherwise predicted to leach. Thorough knowledge and monitoring of the system are essential for such treatment schemes, because reoxidation can result in mobilization.

Summary

Dissolution, precipitation, oxidation, and reduction processes are vital to metal cycling in aerobic and anaerobic surface and subsurface environments. As active agents in cycling, bacteria and fungi are fundamental biotic components of natural biogeochemical cycles for metals and metalloids. Some important processes catalyzed by microorganisms also have important applications in environmental biotechnology in the areas of ore leaching and bioremediation.

Acknowledgments

S.M.G. and T.J.B. acknowledge support from the US Department of Energy Natural and Accelerated Bioremdiation program. G.M.G. gratefully acknowledges financial support from the Biotechnology and Biological Sciences Research Council (94/MAF12243; 94/BRE13640), and British Nuclear Fuels plc. E.P.B. gratefully acknowledges receipt of a Natural Environment Research Council postgraduate studentship.

See also: **Heavy Metals**; **Metal Oxides**

Further Reading

Banfield JH and Nealson KH (eds) (1997) Geomicrobiology: interactions between microbes and minerals. *Reviews in Mineralogy* 35.

Beveridge TJ and Doyle RJ (eds) (1989) *Metal Ions and Bacteria*. New York: John Wiley.

Burford EP, Fomina M, and Gadd GM (2003) Fungal involvement in bioweathering and biotransformation of rocks and minerals. *Mineralogical Magazine* 67: 1127–1155.

Gadd GM (1993) Interactions of fungi with toxic metals. *New Phytologist* 124: 25–60.

Gadd GM (1999) Fungal production of citric and oxalic acid: importance in metal speciation, physiology and biogeochemical processes. *Advances in Microbial Physiology* 41: 47–92.

Gadd GM (2000) Bioremedial potential of microbial mechanisms of metal mobilization and immobilization. *Current Opinion in Biotechnology* 11: 271–279.

Gadd GM (ed.) (2001) *Fungi in Bioremediation*. Cambridge, UK: Cambridge University Press.

Hernandez ME and Newman DK (2001) Extracellular electron transfer. *Cellular and Molecular Life Science* 58: 1562–1571.

Korenevsky A, Glasauer S, and Beveridge TJ (2002) Biomineralization by bacteria. In: Bitton G (ed.) *Encyclopedia of Environmental Microbiology*. New York: John Wiley.

Lovley DR (ed.) (2000) *Environmental Microbe–Metal Interactions*. Washington, DC: ASM Press.

McLean JS, Lee J-U, and Beveridge TJ (2002) Interactions of bacteria and environmental metals, fine-grained mineral development, and bioremediation strategies. In: Huang PM, Bollag J-M, and Senesi N (eds) *Interactions between Soil Particles and Microorganisms*, pp. 228–261. New York: John Wiley.

Verrecchia EP (2000) Fungi and sediments. In: Riding RE and Awramik SM (eds) *Microbial Sediments*, pp. 69–75. Berlin, Germany: Springer-Verlag.

Verrecchia EP and Dumont J-L (1996) A biogeochemical model for chalk alteration by fungi in semiarid environments. *Biogeochemistry* 35: 447–470.

Metals, Heavy *See* Heavy Metals

MICROBIAL PROCESSES

Environmental Factors

P G Hartel, University of Georgia, Athens, GA, USA

Introduction

The activity of all living organisms in soils is affected by environmental factors. These environmental factors are commonly divided into abiotic and biotic factors: the primary abiotic factors include nutrients, water, oxygen, temperature, pH, and light; the primary biotic factors include organism–organism interactions of mutualism, commensalism, amensalism, competition, predation, and parasitism. In addition to these primary factors, other factors (e.g., ionizing radiation in a uranium-rich soil) exist in a few specialized soil habitats, but these rare factors are not considered here.

Before one can predict how an environmental factor will affect the activity of microbes in a particular soil, there are a number of points to consider with respect to the soil, the soil organisms, and the environmental factors. With respect to the soil, it is important to consider how soil differs from other environments. A typical mineral soil is 50% solid material and 50% pore space, and has a bulk density (the weight of the soil divided by the total volume of the soil) ranging from 1.00 to 1.80 $\mathrm{g\,cm^{-3}}$. Because of its density, soil can offer a much broader spectrum of varying environmental factors than other environments. For example, bacteria incapable of tolerating oxygen can exist only micrometers apart from bacteria incapable of existing without oxygen. The creation of these microsites allows soil to contain an exceptional diversity of microorganisms over a small space, and gives soil a complexity unlike all other habitats.

Further, all these organisms live in the soil pores. Under normal circumstances, chemical and physical forces limit movement of most soil organisms in these pores. In terms of chemical forces and bacteria, these organisms are negatively charged at typical soil pH values (5–8) and are held in place by negatively charged clay and organic matter through ion exchange, coordination bonding, hydrogen bonding, and van der Waals forces. In terms of physical forces, soil pores, particularly micropores (less than 30 μm), offer only a tortuous path to movement. Only when soils are saturated, or when plant roots or larger soil organisms actively or passively transport smaller soil organisms, do smaller soil organisms overcome these physical and chemical forces to move great distances (centimeters or more). Therefore, when environmental factors become adverse, most soil organisms are stuck in place and cannot escape to better conditions elsewhere.

With respect to soil organisms, many soil organisms are unidentified. Although the percentage of identification is comparatively good for soil algae and many soil protozoa, nematodes, and other soil fauna, only 0.1–10% of all soil bacteria and archaea, and 5% of all fungi are speciated. Molecular techniques, especially nucleic acid analyses, will increase these percentages, but at present, generalizations about how environmental factors affect microbial activity must be interpreted with caution when the identity of so many species is unknown.

With respect to environmental factors, these factors range from essential to unimportant. Like Goldilocks, soil organisms prefer essential factors to be "just right." For an organism to be outside the boundary of an essential environmental factor means diminished microbial activity, and, if sufficiently far outside the boundary, death (Figure 1). Understanding how environmental factors affect microbial activity is more difficult than it appears, because environmental factors are often dependent on each other.

Abiotic Factors

Nutrients

Nutrients are substances that soil organisms must obtain from the environment to survive and grow. Soils vary in their ability to contain and to hold these nutrients. On a chemical basis, the ability of a soil to hold nutrients is based on its anion exchange capacity and cation exchange capacity, and it is the type and amount of clay and organic matter in soil that determine this.

Some nutrients provide energy, whereas others provide cell constituents. In terms of energy, all soil organisms are either chemotrophs, which derive all their energy from organic and inorganic sources (sometimes also called organotrophs and lithotrophs, respectively), or phototrophs, which convert light energy into chemical energy. Because carbon is the most common element in all soil organisms, chemotrophs and phototrophs are typically further divided into autotrophs, which convert inorganic carbon (i.e., CO_2 or carbonates) into organic compounds, or heterotrophs, most of which are saprophytes and feed on nonliving plants and animals or on humic substances in soil. Table 1 shows examples of this classification.

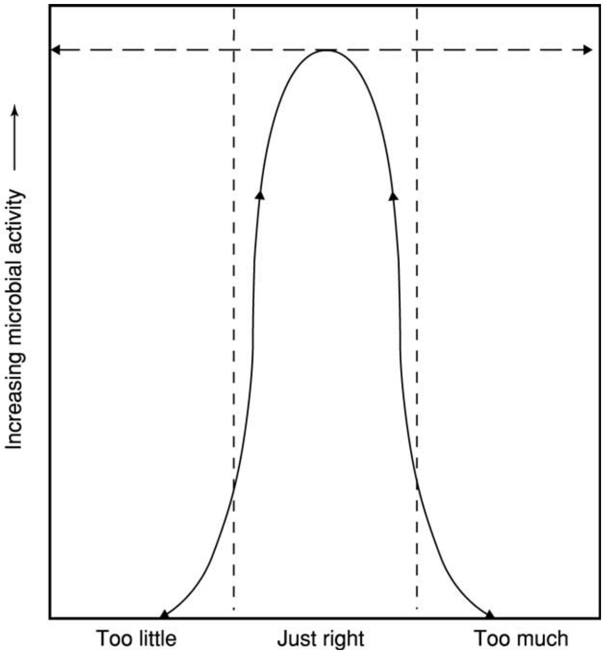


Figure 1 Idealized response of a soil organism to a critical soil environmental factor (solid line). The response of the organism does not have to be bell-shaped, and can be asymptotic to the left or to the right (horizontal dashed line).

In terms of cell constituents, nutrients are divided into macronutrients, substances required in large quantities by organisms, and micronutrients, substances required in small quantities (nanograms per gram of dry weight). Besides carbon, the major macronutrients for soil organisms are nitrogen, phosphorus, and sulfur, and, to a lesser extent, calcium, magnesium, potassium, and sodium. The major micronutrients are cobalt, copper, iron, manganese, molybdenum, and zinc. Macronutrients comprise the vast majority of the structural components such as carbohydrates, proteins, lipids, and nucleic acids, while micronutrients are typically found in growth factors and enzymes.

There are only three natural conditions during which soil organisms see nutrients in abundance. The first is governed by the natural life cycle of plants. As annual plants die and perennial deciduous plants shed their leaves, the accumulated nutrients are released into the soil.

The second condition is the relatively steady exudation of organic compounds from plant roots. Because some soil organisms can promote plant growth through the production of plant growth hormones and other mechanisms, plants have evolved to exude a complex mixture of organic compounds through their roots, primarily organic acids, amino acids, and carbohydrates. The rhizosphere, or zone of soil under the influence of the plant, has sufficient nutrients such that the colony-forming units of bacteria and fungi per gram of rhizosphere soil are typically 10- to 100-fold higher than bulk soil, and the numbers of soil animals that prey on these microbes are doubled or tripled. With such high numbers, it is unsurprising that microbial activity is often intense.

The third condition is when something disrupts the first two conditions. Natural disruptions include conditions such as fire and the episodic drying and rewetting of soil. Fire, by consuming aboveground biomass, minimizes the rhizosphere effect by reducing the energy supply to the soil organisms. Episodic drying causes soil organisms to die, unless these organisms have developed a mechanism to avoid desiccation (e.g., they form a resistant structure such as a spore or cyst). When soil is rewetted, the survivors

Table 1 Classification of all soil organisms based on their source of energy (‘photo-,’ light; ‘chemo-,’ organic or inorganic sources) and carbon (‘auto-,’ CO_2 or carbonates; ‘hetero-,’ organic C sources)

Energy source	*Carbon source*	*Classification*[a]	*Example(s)*
Light	Inorganic carbon	Photoautotroph	Algae, cyanobacteria
	Organic carbon	Photoheterotroph	Photosynthetic bacteria
Organic or inorganic sources	Inorganic carbon	Chemoautotroph	Nitrifying and sulfur-oxidizing bacteria
	Organic carbon	Chemoheterotroph	Fungi, most bacteria, and soil fauna

[a] The suffix ‘-troph’ means to nourish.

dine on the deceased. Human disruptions include conditions such as clear-cutting and tillage. Clear-cutting has the same effect as fire, while tillage disrupts the soil structure, releasing occluded organic matter and causing a temporary glut of nutrients to the soil community.

Bioavailability, or the degree to which nutrients are accessible to soil microorganisms or extracellular enzymes, depends not only on chemical characteristics of the nutrients (e.g., whether they are polar or nonpolar) and their physical place (e.g., whether they are in inaccessible nanopores or not), but also on abiotic factors. For example, in the case of leaf fall in northern temperate regions, microbial activity reaches a peak in autumn, diminishes in winter with its cooler temperatures, and then resurges again in spring when favorable abiotic factors recur. Activity in summer is low because water is often limited.

Not surprisingly, soil organisms have developed an array of different mechanisms to deal with this periodic, episodic, or steady flow of nutrients. Some soil organisms opt for rapid growth only when nutrients are plentiful. By placing all their efforts into reproduction, these organisms attempt to occupy a niche as quickly as possible and then die to low numbers awaiting the next flush of nutrients. Organisms like this are termed ‘r-selected’ or ‘zymogenous’ strategists, and this strategy is characteristic of many bacterial species and soil fauna. Some organisms, by virtue of being left with ‘crumbs from the table,’ opt for a slower growth rate and tend to metabolize more complex nutrients. Organisms like this are termed ‘K-selected’ or ‘autochthonous’ strategists, and this strategy is characteristic of many fungal and actinomycete species. Finally, other organisms lie on the continuum between these two strategies.

For autotrophs, there is no problem of limited carbon, because CO_2 and bicarbonate are usually present in excess of demand. For heterotrophs, microbial activity is dependent on the quality and quantity of available carbon. Readily available natural organic compounds such as simple sugars are degraded quickly and microbial activity is high. For natural organic compounds that are recalcitrant or resistant to degradation (e.g., lignin), microbial activity is low. Toxic compounds such as antibiotics also have the capacity to diminish microbial activity. There is one special case for xenobiotics, anthropogenic compounds that have no biological parallel in the environment. These nonnatural compounds typically have either halogen substituents (e.g., Cl) or are branched polymers (e.g., polypropylene). Soil organisms, by virtue of their enzymatic virtuosity, degrade most xenobiotics as they would natural compounds. Sometimes, however, an organism accidentally metabolizes a xenobiotic in the process of metabolizing a second compound. This process, called cometabolism or fortuitous metabolism, is beneficial to humans in terms of biodegradation of xenobiotics in soils, but is detrimental or at best neutral to the activity of the organism, because the process does not yield any energy or carbon. For example, a bacterium cometabolizing polychlorinated biphenyls (PCBs) yields chlorinated aliphatic and aromatic acids, both dead-end products that the bacterium cannot assimilate.

Water

All soil organisms – without exception – require water for microbial activity. The major force governing water in soil is the total water potential, which is composed primarily of gravitational, matric, and osmotic potentials, and is measured in units of pressure, pascals (Pa). Gravitational potential is the force of gravity pulling water toward the Earth’s center, matric potential is the adsorption of water to surfaces of soil particles and capillary forces, and osmotic potential is the attraction of solute ions for water molecules. Soil organisms have their own water potentials, composed primarily of osmotic and turgor potentials. Before considering these potentials, it is important to consider the relationship among these potentials, soil oxygen, soil texture (the mixture of sand, silt, and clay in soil), and soil porosity. The relationships among temperature, pH, light, and total water potential are considered in later sections.

The major mechanism by which oxygen moves into soil is diffusion. The diffusion coefficient of oxygen in air is $0.189\,cm^2\,s^{-1}$, but the diffusion of oxygen in water is $0.000\,025\,cm^2\,s^{-1}$. Thus, oxygen diffuses 10 000-fold slower in water than in air. How much water a soil holds is dependent on soil texture and soil porosity. Sandy soils have less pore space (35–50%) than clayey soils (40–60%), but the pore space in sandy soils is mostly macropores (30–100 μm), whereas the pore space in clayey soils is mostly micropores (less than 30 μm). For this reason, sandy soils drain quickly and hold less water than clayey soils, which drain more slowly and hold more water.

The theoretical limits for microbial activity in soil are between 0 and -81 MPa. At 0 MPa, soil is completely saturated; at -81 MPa, DNA is denatured. Besides the effect of each water potential on microbial activity, it is also important to consider oxygen, movement, and nutrients. When the soil is saturated, diffusion of oxygen into the soil is slowed, and if the oxygen consumption rate exceeds the diffusion rate, the soil, regardless of its texture and porosity, becomes anaerobic. Heterotrophs incapable of anaerobic respiration (e.g., most fungi) fare poorly. Heterotrophs capable of anaerobic respiration (e.g., many bacteria) fare better but show diminished activity because the amount of energy produced under anaerobic conditions is less than under aerobic conditions. Phototrophs on the soil surface (e.g., eukaryotic algae and cyanobacteria) fare well and even prefer saturated conditions. Movement by aquatic soil organisms is maximal because the water film thickness around soil particles is maximal. Texture plays a role because soil organisms move more easily in sandy soils than in clayey soils (again, because sandy soils have more macropores). Concentrations of water-soluble nutrients are low. Water stress for almost all soil organisms is minimal.

Once water drains from the macropores, the water potential becomes negative (approximately −0.01 MPa for sandy soils and −0.03 MPa for other soils), and oxygen diffusion increases. Sandy soils, which contain more macropores than clayey soils, are more oxygenated than clayey soils. Except for movement by aerial (i.e., living in air-filled spaces) soil animals such as microarthropods (mites and collembolans), movement by aquatic soil organisms such as nematodes, protozoa, oomycete zoospores, and bacteria becomes more difficult as the water-film thickness dwindles to only a few micrometers. However, the combination of aerobic conditions, concentrated nutrients, and good water relations means microbial activity is near maximal.

As the total soil water potential decreases from −0.01 to −0.1 MPa, oxygen diffusion improves, but water films become discontinuous, and consequently nutrient availability and movement by aquatic organisms decrease. Because matric potential is related to capillary forces and these forces are directly related to the soil pore diameter, clayey soils, by virtue of their more numerous micropores, contain more water than sandy soils and dry more slowly. This drying rate is important, because the faster the rate of drying, the greater the mortality of the organism. Because microbial activity depends on each soil organism maintaining proper turgor pressure for cell functioning, soil matric and osmotic potentials increase in their importance. Except in saline soils, matric potentials are more important than osmotic potentials.

By −1.5 MPa, water is gone from the micropores and only water bound to soil solids is available for microbial activity. The soils are relatively oxygenated; movement stops. Diffusion of nutrients is so retarded that starvation conditions are created. Furthermore, because all soil organisms contain semipermeable membranes that allow water to enter and exit in near equilibrium with the environment, osmotic potential becomes critical as soil ions become more and more concentrated. To avoid plasmolysis (contraction of the cytoplasm from the cell wall), most soil organisms divert their energy to accumulating stress solutes (e.g., K^+) and compatible solutes (e.g., glycerol; so named because they must be compatible with cell functioning) to counter this negative potential. By −4.0 MPa, most biological processes (e.g., nitrification) stop and only variously named osmophiles, halophiles, or xerophiles show appreciable microbial activity. Unless an organism can form a resistant structure (e.g., spores in bacteria and fungi, akinetes in the cyanobacteria, cysts in the protozoa), most organisms desiccate and die.

Oxygen

As already noted, oxygen in soil depends on soil water, soil texture, and soil porosity. Both soil depth and temperature can be added to this mix: generally, oxygen decreases and CO_2 increases with increasing depth, and gas solubility, particularly in water, decreases with increasing temperature. Most soil organisms are obligate aerobes and use oxygen exclusively as a terminal electron acceptor. Because aerobic respiration produces more energy per mole of substrate than anaerobic metabolism, obligate aerobes recover more energy for biomass production than anaerobic organisms. For this reason, population densities of obligate aerobes are higher than those of anaerobes. Without a source of oxygen, aerobic respiration ceases and no energy is produced. In contrast, obligate anaerobes cannot survive the presence of oxygen because aerobic respiration temporarily yields hydrogen peroxide (H_2O_2) and superoxide (O_2^-), toxic products to which obligate anaerobes lack the enzymes (catalase and superoxide dismutase, respectively) to break down. As a result, obligate anaerobes live strictly by fermentation. Some obligate aerobes that prefer oxygen as the terminal electron acceptor can respire anaerobically by substituting nitrate (NO_3^-), sulfate (SO_4^{-2}), or carbonate (CO_3^{-2}) for oxygen. This anaerobic respiration is restricted to bacteria and for this reason these bacteria dominate in anaerobic soils. Many bacteria and some fungi are facultative anaerobes, which grow in the presence or absence of oxygen. When oxygen is available, the organisms grow aerobically; when oxygen is unavailable, the organisms grow by either fermentation or anaerobic respiration. Finally, aerotolerant anaerobes lack an electron transport system and function solely by fermentation.

Redox (reduction–oxidation) potential, the tendency of a compound to accept or donate electrons, can also give some insight to oxygen tensions in soil and which organisms are likely to be active in a particular soil environment. From a standpoint of processes, soils with a positive redox potential (0–400 mV) are ‘oxidizing’ habitats and are conducive to aerobic processes, whereas soils with a negative redox potential (0 to −400 mV) are 'reducing' habitats and conducive to anaerobic processes such as methane production. However, it would be a misstatement that highly aerobic soils contain few or no anaerobes. The ability of soils to contain water-filled micropores restricts oxygen diffusion and guarantees that anaerobic microsites exist. Thus, it is common to find obligate anaerobes in the upper soil horizons. Also, some anaerobic processes, such as denitrification, can occur in soils with a positive redox potential when molecular oxygen is absent but other terminal electron acceptors are present.

Temperature

For soil organisms, it is important to consider not only the temperature at which microbial activity is optimal, but also the minimum and maximum temperature that the organism can tolerate. Soil organisms are divided into psychrophiles, mesophiles, and thermophiles. Psychrophiles do not grow above 20°C, have minima of 0°C or lower, and optima around 16°C. Mesophiles have minima of above 0°C, maxima below 50°C, and optima between 15 and 40°C. Finally, thermophiles have minima at or above 20°C, maxima at or above 50°C, and optima between 40 and 50°C.

There is an important relationship between temperature, water potential, and microbial activity. Both water potential and temperature must be within critical range for microbial activity to occur. Assuming water potential is within this range, at one extreme, reduced temperatures slow cellular processes. At the other extreme, excessive temperatures denature proteins and alter cell membrane permeability. At either extreme, soil organisms typically alter the fatty acyl composition of their membrane lipids and produce either cold-shock or heat-shock proteins to withstand these stresses. Between these extremes is another relationship, often referred to as the 'Q_{10} rule for biological systems,' where Q_{10}, the temperature quotient, equals 2. This rule states that, within a limited range, microbial activity doubles for each 10°C increase in temperature. For example, mesophilic organisms at 30°C will have twice the activity as the same organisms at 20°C. Again, this rule is a range, and there are examples, such as petroleum mineralization in Alaska, where Q_{10} values much greater than 2 have been measured.

Because most soil temperatures correspond well with the temperature optima of mesophiles, it is not surprising that most soil organisms are mesophiles. However, some specialized habitats (e.g., volcanic soils) do favor thermophilic actinomycetes and fungi.

pH

As the negative log of the hydrogen ion concentration, pH is intimately connected to water in the soil and the soil organisms. With regard to soil, wherever rainfall is abundant, water leaches bases from the soil and the soils are predominantly acid; wherever rainfall is scant, there is not enough water to leach the bases from the soil and the soils are predominantly alkaline. Because soils have a pH-dependent charge originating from organic matter and the broken edges of clays, this charge decreases as the soil becomes more acid, reducing not only the cation exchange capacity of the soil, but also the ability of the soil to retain nutrients. Reducing the pH does not necessarily reduce the entire cation exchange capacity in soil; some charge may be associated with isomorphous substitution in clay minerals and this charge is pH-independent. With regard to soil organisms, when soils are saturated, the hydrogen ion concentration is diluted; when soils are dry, the hydrogen ion concentration is concentrated and this concentration affects the ability of soil organisms to control anions and cations and offset high osmotic potential.

Although some bacteria and soil fauna tolerate low pH (acidophiles; able to grow at pH 3 or lower) or high pH (alkaliphiles; able to grow at pH 9 or higher), most bacteria and soil fauna are neutrophiles and prefer a near-neutral pH range of 6–7. The same is true of soil algae, which again are found in acid and highly alkaline environments, but prefer pH ranges of nearly neutral to slightly alkaline. Many bacteria and algae prefer this pH range because soil nutrients are most available. Furthermore, a neutral pH avoids heavy metal toxicity, because almost all these metals (a density of greater than 5 Mg m^{-3}) are only active in acid soils (pH <5). In contrast, fungi are acid-tolerant (pH 4–7) and tend to be abundant in acid soils, where there is less competition from bacteria.

Light

For phototrophs such as cyanobacteria and algae, light is an essential factor because it is required as a source of energy. Soil affects light because soil constituents absorb or reflect it (the albedo effect), and soil disturbances (e.g., tillage) affect how far light penetrates into the soil. For these reasons, the photic zone in soil is variable. Nevertheless, phototrophs decrease with increasing soil depth and decreasing light intensity. Except for a few phototrophs that are facultatively capable of heterotrophy, phototrophs deep in the soil are likely to be metabolically inactive.

For chemotrophs, light is important for a different reason. Although visible light may be important for certain phototrophic responses (e.g., spore formation in fungi), ultraviolet-B (UV-B) light (290–320 nm) is a source of nonionizing radiation. This irradiation splits water molecules into free radicals that damage DNA. This irradiation links light to water potential: the higher the water content, the more free radicals produced, and the higher the DNA damage. Although ozone screens out most UV-B light, ozone depletion in the atmosphere offsets this. Therefore, it is likely that this abiotic factor will increase its deleterious effect on microbial activity in the photic zone, and only soil organisms that produce UV-absorbing pigments or develop rapid mechanisms of DNA repair will be relatively unaffected.

Biotic Factors

Soil organisms can also interact with other organisms, and, with the exception of neutralism (no interaction), these interactions affect their microbial activity. These biotic interactions can also include plants and larger soil fauna, and at their most complex, populations of organisms or consortia. These interactions are mutualism, symbiosis, commensalism, competition, amensalism, predation, and parasitism (Table 2).

Symbiosis or Mutualism

In symbiosis or mutualism, both organisms benefit. This association can be specific or general and, in some cases, so highly specialized that one or both organisms are obligate symbionts and cannot live without the other. Relationships where there is mutual benefit but no obligation on the part of either organism are often termed ‘synergistic.’ Because of its importance to agriculture, the best-known symbiosis is the presence of nitrogen-fixing bacteria in plants. In return for the bacteria fixing atmospheric nitrogen, the plant supplies the bacteria with nutrients. An agricultural example is the symbiosis between *Bradyrhizobium japonicum* and the legume soybean (*Glycine max*); a forestry example is the symbiosis between the actinomycete *Frankia* and a range of angiosperms such as alder (*Alnus* sp.). A similar situation is observed with several genera of soil fungi, the mycorrhizae, which colonize plant roots, thereby enhancing uptake of water and mineral nutrients by the plant, principally phosphorus. Again, the plant supplies the mycorrhizae with nutrients. Symbiosis or mutualism is much more frequent than commonly realized. For example, 95% of all plants are mycorrhizal. These kinds of mutually cooperative ventures – not competitive ones – are the norm in nature.

Table 2 All possible positive interactions (+), no effect (0), and negative interactions (−) between two organisms. The effect of these interactions on microbial activity corresponds with the symbol

Organism No. 1	*Organism No. 2*		
	+	*0*	*−*
+	Mutualism, symbiosis	Commensalism	Predation, parasitism
0	Commensalism	Neutralism	Amensalism
−	Predation, parasitism	Amensalism	Competition

Commensalism and Amensalism

In commensalism, one organism benefits while the other is unaffected. For example, one organism can provide an essential growth factor, such as a vitamin, for another organism. This type of cross-feeding is common in soil organisms. The opposite of commensalism is amensalism, where one organism is harmed while the other is unaffected. A good example of this interaction is when one organism produces an antibiotic against another organism. Such an interaction is often the basis of biological control. For example, some isolates of the bacterium *Pseudomonas fluorescens* can suppress the fungal pathogen *Gaeumannomyces graminis*, responsible for ‘take-all’ in wheat (*Triticum aestivum*).

Competition

In competition, both organisms vie for the same resource. This resource can be for any one of the abiotic factors as well as space. In the case of nutrients, the three factors that govern competitive success are growth rate, substrate affinity, and efficiency. If nutrients are unlimited, then microbial activity is governed by the maximal growth rate. This maximal growth rate is rarely the case in soil; the norm for most organisms in soil is starvation. Under these circumstances, substrate affinity and efficiency govern the ability to compete. In the case of substrate affinity, assuming an organism can metabolize a substrate, an organism that recognizes a substrate at a low concentration outcompetes an organism that only recognizes a substrate at a high concentration. In the case of efficiency, an organism that is more efficient in its carbon and energy use outcompetes a less efficient organism.

Predation and Parasitism

Finally, in predation and parasitism, one organism feeds on another. The most common example of predation in soil is protozoa and nematodes preying on bacteria and fungi, particularly in the rhizosphere, where numbers of bacteria and fungi are high. In Figure 2, two types of soil protozoa, a ciliate (*Colpoda* sp.) and a flagellate (*Bodo* sp.), as well as a bacteriovorous nematode, are preying on bacteria attached to the clay particles and organic matter, while, above, a mycophagous mite is preparing to eat some mycorrhizal hyphae. Predator responses follow classic predator–prey curves: increases and decreases in numbers of predators lag behind increases and decreases in numbers of prey, respectively. Numbers may not be a good reflection of microbial activity, however. The C:N ratio of soil fauna is approximately 20:1, whereas the C:N ratio ranges from 4 to 5:1 for bacteria and up to 15:1 for fungi. When protozoa prey on bacteria and fungi, nitrogen is in excess and is excreted into the environment. This nitrogen is then available to promote microbial activity.

The most common example of parasitism in soil is viruses. Even though viruses are nonliving (because they have no intrinsic metabolism or ability to replicate), they parasitize numerous soil bacteria, fungi, and soil fauna. This parasitism is normally a negative interaction, but viruses can sometimes interact positively because they are capable of transduction (adding a new gene or modifying an existing gene in the host DNA). If this addition or modification makes the host more competitive, then the change benefits microbial activity.

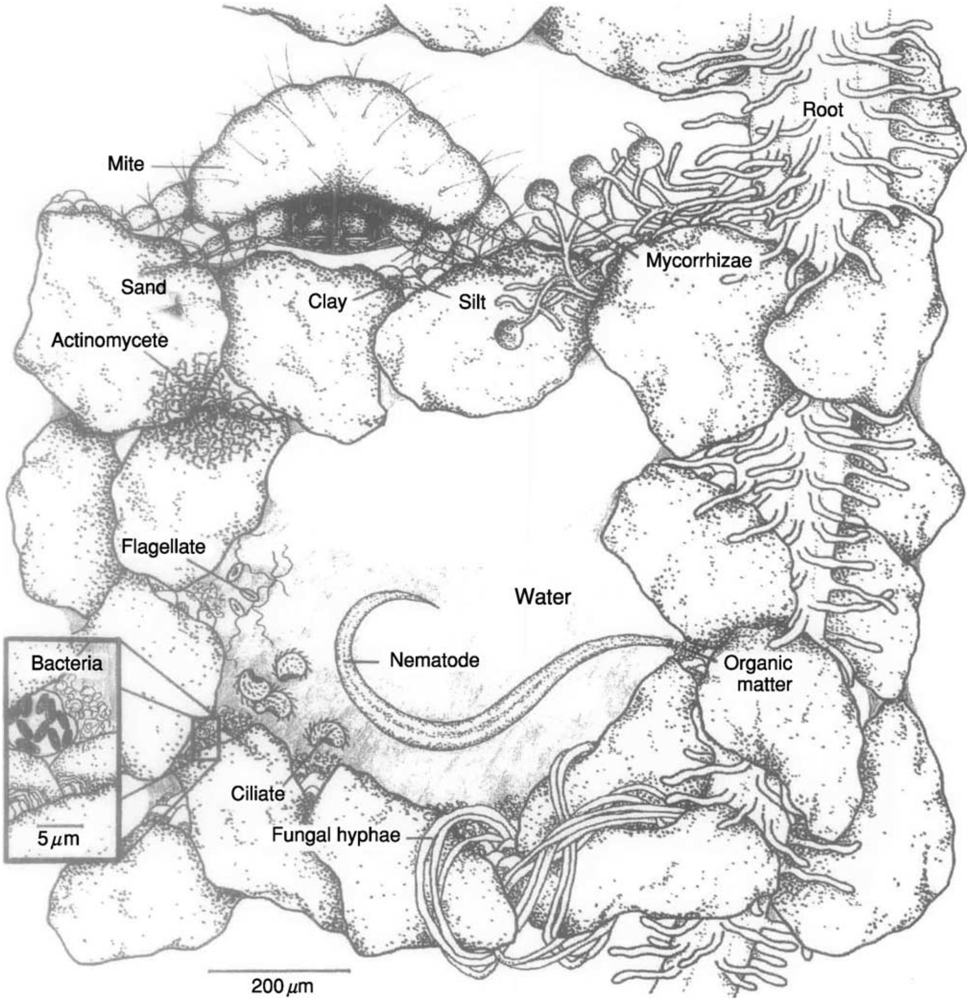


Figure 2 A sample of well-aggregated soil, slightly greater than 1 mm^2, showing a variety of biotic and abiotic factors. As gravitational water flows out of the two largest pores, water potential and oxygen vary from top to bottom, while a small root creates a rhizosphere effect, a gradient of nutrients from high (right) to low (left). Root transpiration affects the water potential depending on whether it is day (more) or night (less). The root alters the soil pH by releasing H^+, HCO_3^-, and organic compounds, depending on the type of nitrogen in the soil (NO_3^- increases, NH_4^+ decreases), symbiotic relationships (here, mycorrhizae), and microbial activity. Temperature also affects the rate of microbial activity. Biotic factors, and interactions of biotic and abiotic factors are noted in the text. In all, a wide variety of microsites are being created, many only micrometers apart. (The mite has been turned head-on to fit.) (An original drawing by Kim Luoma.)

Integrating Abiotic and Biotic Factors

In addition to interactions between and among soil abiotic factors, and interactions between soil organisms, there are also interactions between the abiotic and biotic factors. In one direction, soil abiotic factors affect biotic interactions. For example, nutrient availability, water potential, and oxygen profoundly affect microbial interactions. In Figure 2, the aerial soil mite would not be preparing to eat the mycorrhizae unless the limited soil nutrients (e.g., phosphorus) had promoted the symbiotic relationship between the plant and the fungus, and the gravitational water had drained from the pore, permitting sufficient oxygen diffusion for the mite to breathe. In the other direction, biotic interactions affect soil abiotic factors, both physically and chemically. In Figure 2, actinomycetous and fungal hyphae, and bacteria in a extracellular polysaccharide 'plug' are binding the soil particles, thus promoting soil aggregation, one of the most important physical effects of microorganisms on soil. This aggregation affects soil porosity, which, in turn, affects soil water and oxygen relations. As an example of a chemical effect, the nematode and protozoa eat the bacteria and excrete excess ammonia, which chemoautotrophic bacteria convert to nitrite, which, in turn, is converted commensalistically by nitrite-oxidizing bacteria to nitrate. These acidifying reactions lower the soil pH.

The interactions between and among abiotic and biotic factors are the equivalent of an intricate and elegant dance with many partners. Our understanding of this dance, particularly among microbial communities, is still rudimentary. The major barriers to understanding this dance are our poor understanding of the identity of most soil archaea, bacteria, fungi, and undoubtedly many of the soil fauna, and the nature of the soil matrix to provide a multitude of microsites, each with their properties, over distances that may be only a few micrometers apart. This is what gives soil a spatial and temporal dynamicism and heterogeneity unlike all other habitats.

See also: **Microbial Processes:** Community Analysis; Kinetics

Further Reading

Brady NC and Weil RR (2004) *Elements of the Nature and Properties of Soils*, 2nd edn. Upper Saddle River, NJ: Pearson/Prentice–Hall.

Coleman DC and Crossley DA Jr (1996) *Fundamentals of Soil Ecology*. New York: Academic Press.

Coyne M (1999) *Soil Microbiology: An Exploratory Approach*. Albany, NY: Delmar Publishers.

Darbyshire JF (1994) *Soil Protozoa*. Wallingford, UK: CAB International.

Havlin J, Tisdale SL, and Beaton JD (1998) *Soil Fertility and Fertilizers: An Introduction to Nutrient Management*, 6th edn. New York: Macmillan.

Horikoshi K and Grant WD (1998) *Extremophiles: Microbial Life in Extreme Environments*. New York: Wiley–Liss.

Madigan MT, Martinko JM, and Parker J (2000) *Brock's Biology of Microorganisms*, 9th edn. Upper Saddle River, NJ: Prentice-Hall.

Nester EW, Roberts CE, and Nester MT (1998) *Microbiology: A Human Perspective*. New York: WCB/McGraw-Hill.

Parr JF, Gardner WR, and Elliott LF (1981) *Water Potential Relations in Soil Microbiology.* SSSA Special Publication No. 9. Madison, WI: Soil Science Society of America.

Sylvia DM, Fuhrmann JJ, Hartel PG, and Zuberer DA (1998) *Principles and Applications of Soil Microbiology.* Upper Saddle River, NJ: Prentice-Hall.

Torsvik V, Sørheim R, and Goksøyr (1996) Total bacterial diversity in soil and sediment communities – a review. *Journal of Industrial Microbiology* 17: 170–178.

Wicklow DT and Söderström B (1997) *Environmental and Microbial Relationships*, vol. 4. In: Esser K and Lemke PA (eds) *The Mycota: A Comprehensive Treatise on Fungi as Experimental Systems for Basic and Applied Research*. Berlin, Germany: Springer-Verlag.

Community Analysis

C H Nakatsu, Purdue University, West Lafayette, IN, USA

Introduction

The ecological diversity of soil microbial communities is a measure of microbial community richness and abundance. Richness is an indication of the number of different species present, whereas abundance is the number or biomass of each species present. Microbial communities can also be analyzed for their functional role in soil. There is not a single method available that can provide complete information on a microbial community; the method chosen is dependent on the question being posed. In the past, soil scientists relied mainly on cultivation and measuring substrate transformation to assess and monitor microbial communities. Now a common view in soil microbiology is that cultivation-dependent methods represent only a small fraction of the microbial community present in soil. Beginning in the 1980s, Vigdis Torsvik and her advisor, Professor Jostein Goksøyr, began testing the use of deoxyribonucleic acid (DNA) reassociation kinetics to demonstrate the true extent of soil microbial richness. They have concluded that 1 g of soil has an estimated 4000–10 000 different bacterial genomes present, of which less than 1% can be readily cultivated. This work has prompted largely the move from isolation and phenotypic characterization of microbial communities to direct analysis using genetic and biochemical tests to survey a more representative group. Using a variety of molecular techniques, investigators are studying soil microbial communities from a wide range of habitats and geographic locations. Their results generally demonstrate that soil environments are far more diverse, even under harsh environmental conditions, than was expected and that previously undescribed groups are important components of the soil community. Analytical approaches adopted are crucially dependent on the questions being addressed and availability of equipment and resources.

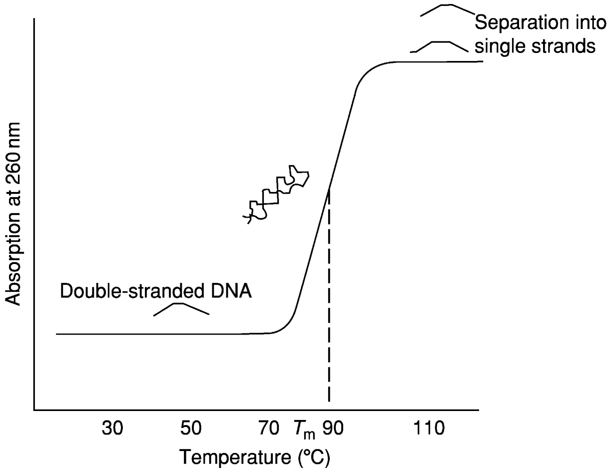


Figure 1 Melting curve of DNA showing the melting temperature, T_m, when half the DNA is denatured.

Background

Nucleic Acids

Nucleic acids are polymers of nucleotides joined by covalent (phosphodiester) bonds formed between their sugar and phosphate groups. There are two types of nucleic acids that differ chemically because of their sugar component, deoxyribose in DNA and ribose in ribonucleic acid (RNA). Physically the two differ because DNA is composed of two strands that form a helix and RNA is single-stranded. Attached to the sugar and phosphate groups are nucleotide bases of which there is a total possibility of five: cytosine (C), thymine (T, in DNA only), uracil (U, in RNA only), adenine (A), and guanine (G). The two strands comprising DNA are joined by hydrogen bonds between complementary nucleotide bases, where G always bonds with C (by three hydrogen bonds) and A with T (by two hydrogen bonds). The relative content of G + C compared with A + T in the genome of an organism can be used to differentiate taxonomic groups. The different order or sequence of nucleotide bases codes the information to determine appearances and functions in every organism. Nucleotide sequences that code for proteins are called genes. After nucleotide sequences are obtained, one can compare them with sequences in a databank (e.g., GenBank, EMBL) to determine whether any similar sequences have been found. This can provide information about potential function and phylogeny of a gene.

Denaturation and Renaturation

Denaturation of DNA occurs when the weak hydrogen bonds between the double strands are disrupted and the molecule becomes single-stranded. Thus the rate of denaturation is dependent on the proportion of G + C versus A + T bases. This process can be reversed in a process called renaturation or annealing. In the laboratory, DNA can be denatured by applying heat, increasing pH, or by adding denaturing chemicals (e.g., urea, detergents). In addition to these factors, viscosity and ionic strength can be adjusted to influence rates of denaturation and renaturation. As the DNA strands become denatured, different measurable properties of the molecule changes over time. Often DNA is quantified spectrophotometrically by measuring its ultraviolet (UV) light absorption at 260 nm. This approach can be used to track the melting characteristics of DNA, because in the single-stranded form it absorbs more light than when it is double-stranded. The data when graphed reveal the melting curve of the DNA (Figure 1). The melting temperature, T_m, is the transition temperature at which half the strand is denatured, as reflected in the halfway point in the change in absorbance at 260 nm. This temperature is important when carrying out experiments such as hybridization and polymerase chain reaction (PCR), where denaturation and/or renaturation of DNA strands is required.

Soil Nucleic Acid Extraction

Nucleic acids can be extracted directly from soil to analyze the microbial community. Alternatively cells can first be separated from the soil then the nucleic acids extracted. The latter approach decreases contamination by soil components but may also be biased toward cells that are separated more easily from soil. Most researchers isolate DNA from soil because it is more stable than RNA; however, protocols have been developed for soil RNA isolation. All methods for nucleic acid extraction and purification follow the same principles. Cell lysis is the first step, followed by separation of the nucleic acids from the other cellular components, and finally concentrating the nucleic acids into a working volume. There are many papers that outline these methods, and commercial kits are marketed by a number of companies. An important factor when extracting nucleic acids from soil is to ensure the approach achieves equivalent lysis of all cell types so that the analysis is representative of the community. This is the main source of bias in direct-analysis methods of soil-extracted nucleic acids. Many laboratories have chosen to use mechanical methods such as bead-beating to ensure cell lysis. Lysis efficiency can be determined by microscopic examination of the sample before and after the extraction procedure. The other factor that must be considered is soil composition, because nucleic acids are charged and will bind to components, particularly clay. After the nucleic acids are extracted, a number of methods can be considered that examine nucleic acids directly or indirectly after amplification by PCR.

Ribosomal RNA

Ribosomes are comprised of proteins and ribosomal RNA (rRNA) and are found in all organisms, because they are the sites for protein synthesis. Presently, the genes coding for rRNA are used most extensively to obtain phylogenetic information. These genes were chosen because: they are present in all organisms; rRNA function is conserved; rRNA sequences have regions with high and low evolutionary conservation, making them phylogenetically informative; and, in actively growing cells, ribosomes can occur in high copy numbers (more than 10 000). The ribosome is composed of two subunits, referred to as small and large rRNA, differentiated by their sedimentation coefficient, *S*. In prokaryotes the small subunit contains 16*S* rRNA and the 70*S* large subunit contains both 23*S* and 5*S* rRNA. In eukaryotes the small subunit is slightly larger and contains 18*S* rRNA and the 80*S* large subunit contains 28*S*, 5.8*S*, and 5*S* rRNA. The 16*S* rRNA gene has been used the most in prokaryotic soil-community analysis. Many laboratories are determining 16*S* rRNA sequences from environmental sources and adding them to databanks, increasing the utility of using these genes for community analysis. In eukaryotic organisms the sequences between the rRNA genes, i.e., the intergenic region, are often included in analyses, because the gene sequence itself is often too conserved to differentiate species. The rRNA or rDNA can be isolated directly from soil and examined directly or after amplification using PCR. Many of the soil-community analytical techniques discussed herein are based on analysis of these genes or gene regions. It is important to note that the number of rRNA gene copies in an organism is variable, therefore one must not assume that determining the number of rDNA copies is equivalent to counting the number of cells in a community.

Direct Community Analysis

G + C Content

The G + C method is a coarse scale method, because different populations with similar G + C content will group together. It is used to determine compositional differences between microbial communities in different soils or after the soils have been perturbed, for example by the addition of a pesticide. A dye that preferentially binds to A + T or G + C nucleotides is added to DNA extracted from a soil community. Presently, the most commonly used dye is bisbenzimidazole (Hoechst 33258), which binds to A + T-rich regions. The buoyant density of the DNA decreases as more dye is bound to AT regions. Populations with different A + T content are then separated by cesium chloride density gradient centrifugation. Populations with similar G + C content will layer at the same gradient concentration. These layers can be collected in fractions and quantified by spectrophotometry to estimate G + C group sizes. Each of these fractions can be differentiated again by density gradient centrifugation at a finer scale, or some of the methods described below can be used to obtain more information about community composition. This method is not recommended if large quantities of DNA (more than 50 μg) and an ultracentrifuge are not available. The number of samples examined can be limited by the capacity of sample handling in the laboratory and amount of soil available for DNA extraction.

DNA Reassociation

A method to determine species richness in a soil community is reassociation of DNA strands after they have been completely denatured. It is based on the assumption that the rate of reassociation is dependent on the number of different genomes in a sample. Strand reassociation takes longer when the genetic heterogeneity of a community is greater. This method can be very time-consuming, because soils of high heterogeneity can take weeks to reassociate. The heterogeneity of the sample is proportional to the $\mathrm{Co}t_{1/2}$ value, where Co is the initial molar concentration of DNA and $t_{1/2}$ is half the time required for complete strand reassociation. The number of different species can then be estimated by comparing the sample $\mathrm{Co}t_{1/2}$ value to that of a reference strain with a known genome size. This method requires large quantities

(25 μg) of highly purified DNA, sheared to equivalent lengths. Due to the length of time required for reassociation, it is mandatory to have a UV spectrophotometer equipped with a thermally controlled sample cell dedicated to the analysis.

Nucleic Acid Hybridization

Nucleic acid hybridization methods can be used in soil community analysis to assess the composition, physical distribution, abundance, and/or relatedness of nucleotide sequences to DNA or RNA extracted from soil samples or left intact in cells (discussed in the next section). Nucleic acid hybridization methods rely on the ability of complementary single-stranded nucleotide sequences to anneal to form nucleic acid duplexes (discussed in the section Denaturation and Renaturation). This is commonly accomplished by hybridization (or annealing) of a previously characterized nucleic acid sequence (probe) to its counterpart (target) within a mixture of DNA or RNA (Figure 2). Probes are labeled with a radioisotope or fluorescent marker for detection. The target is

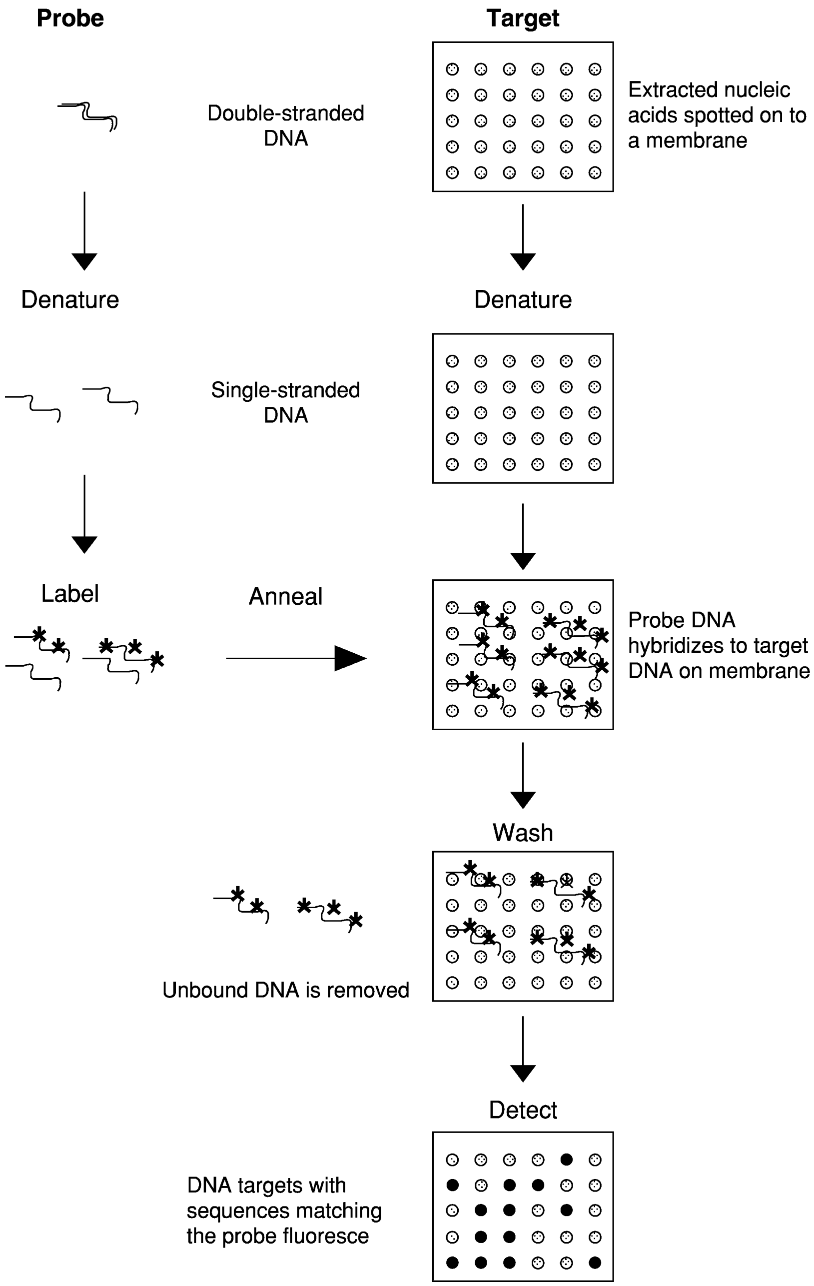


Figure 2 Hybridization and detection of a fluorescently labeled probe to target DNA that has been attached to a nylon membrane.

immobilized on to a solid support such as a nylon membrane or glass slide to facilitate hybrid detection. For these assessments, the nucleic acid sequences need not match exactly, because a limited number of base-pair mismatches can be tolerated if hybridization conditions (temperature, salt concentrations) are properly manipulated. Quantification of the active community members can be accomplished by using rRNA as a target instead of DNA, because the rRNA numbers increase with cell activity. Microarray technology is now making it possible simultaneously to hybridize to thousands of targets on a relatively small surface area. If adequate quantities of target nucleic acids are not readily obtained by extraction, potential targets can first also be amplified by PCR.

To obtain a relatively coarse measure of community composition, 16*S* rRNA gene probes are often used. Probe sequences have been determined that can differentiate community members at phylogenetic levels, ranging from broad (all organisms, kingdom, classes) to narrow (species or strain; examples given in Table 1). The broad probes can provide numbers of different groups present in a community. It is also possible to use all these sequences as targets for microarray analysis. However, due to the large number of microbial species still not described from soil, it is presently not possible to use this method to determine species richness. The species- and strain-specific probes are also useful to track the fate of a specific microorganism within a soil community. This approach has been used extensively to detect the presence of specific functions or functional groups such as ammonia-oxidizing bacteria in a soil community. If microarrays are not being used, then this method does not require specialized equipment other than those found in standard molecular biology laboratories.

Microscopic Examination and Fluorescent *In Situ* Hybridization

To examine microorganisms in the context of their soil environment, analysis can be made microscopically. An *in situ* hybridization-based procedure is employed in which fluorescently labeled probes diffuse into the cell, hybridize to the target rRNA, and can then be observed microscopically. Microscopic observations of bacterial cells are typically made using phase-contrast, epifluorescent, or confocal laser microscopes. Cells are usually killed then stained to detect all organisms or specifically to detect certain groups. General stains such as acridine orange, Hoechst 33258, and ethidium bromide bind to nucleic acids and thus can be used to detect all organisms. Specific groups are detected using the rRNA as the target, because this target has a high copy number in active cells and it is single-stranded to allow hybridization to take place. Sequences have been determined that are specific to different phylogenetic levels, ranging from all organisms to specific species. The availability of different-colored fluorescent markers makes it possible simultaneously to observe cells hybridized to different probes. In addition to determining the spatial distribution of cells in a community, this approach is beneficial because it is cultivation-independent and does not suffer from the biases of PCR. However, as in all methods, there are other shortfalls that must be considered when deciding the appropriateness of its use. Some of the limitations are: background fluorescence from the soil; interference by soil components; low cellular ribosome content making it difficult to see labeled cells; poor penetration of probe into the cell; and differential penetration of probe into different cell types.

Table 1 Examples of oligonucleotide probes targeting Bacteria at different phylogenetic levels

Probe name	*Target organism*	*Sequence*
UNIV1390	All organisms	5′ GAC GGG CGG TGT GTA CAA 3′
EUB338	Domain Bacteria	5′ GCT GCC TCC CGT AGG AGT 3
PLA46	Planctomycetes	5′ GAC TTG CAT GCC TAA TCC 3′
ALF968	α-Proteobacteria	5′ GGT AAG GTT CTG CGC GTT 3′
BET42*	β-Proteobacteria	5′ GCC TTC CCA CTT CGT TT 3′
GAM42*	γ-Proteobacteria	5′ GCC TTC CCA CAT CGT TT 3′
ACID20	Acidobacteria	5′ CAG CGT TGA TTC TGA GCC A 3′
GPOS1192	Gram-positive bacteria	5′ TAA GGG GCA TGA TGA TTT GAC G 3′
HGC69*	Actinobacteria	5′ TAT AGT TAC CAC CGC CGT 3′
LGC353	Low G + C Gram-positive	5′ AAG ATT CCC TAC TGC T 3′
BACIL1136	Bacillus	5′ AGT CAC CTT AGA GTG CCC AAC T 3′

All probes target the 16*S* rRNA with the exception of probes with * designations, which target 23*S* rRNA.

Source: El Fantroussi S, Urakawa H, Bernhard AE, Kelly JJ, Noble PA, Smidt H, Yershov GM, and Stahl DA (2003) Direct profiling of environmental microbial populations by thermal dissociation analysis of native rRNAs hybridized to oligonucleotide microarrays. *Applied and Environmental Microbiology* 69: 2377–2382.

Fatty Acid Analysis

Although many researchers are using nucleic acid-based analyses, phospholipid fatty acids (PLFA) have also been used extensively as a cultivation-independent method for soil-community analysis. With the exception of Archaea, phospholipids are found in cell membranes of all living organisms and are rapidly degraded upon their death. Thus this analysis can provide a measure of viable community biomass, structure, and at a crude level, composition. Unlike most of the other methods discussed that are genotypic, this is a phenotypic approach. There are many different PLFA that can be identified by gas chromatography after their conversion to fatty acid methyl esters (FAME). The PLFA profile of a microorganism is phylogenetically informative. The mixture of different PLFA profiles extracted from a soil can be used as a fingerprint of the community structure. Changes in community structure can be monitored by following changes in their PLFA fingerprint profiles. There are 'signature' PLFA profiles found in certain groups, which allow the identification of the presence and quantity of PLFA in a community. For example, there are fatty acids specific to fungi, Gram-positive bacteria, Gram-negative bacteria, and plants, which can then be used to calculate diversity of the community at a coarse level. However, there are overlapping fatty acids between microorganisms in a community, therefore this method cannot be used to identify or track specific species. Also, due to the ubiquity of lipids, it is essential that great care is taken during sample handling and processing in the laboratory. The method requires an evaporator to dry samples for storage, a spectrophotometer for biomass measures, and a gas chromatograph for FAME analysis.

Indirect Community Analysis

Polymerase Chain Reaction

PCR is extremely useful in soil-community analysis, because an organism does not need to be cultivated for the procedure to be used. It mechanizes the steps in nucleic acid replication in the laboratory, thus allowing the amplification of a specific nucleotide sequence by approximately 10^6-fold. DNA replication is semiconservative. Each strand of DNA serves as a template, and the newly replicated daughter strand is complementary to the parent strand because of basepair specificity. The nucleotides are added to the growing strand by an enzyme called DNA polymerase. Polymerization begins at a primer site and always occurs by joining the carbon 5 ($5'$) of one sugar to carbon 3 ($3'$) of the adjacent sugar ($5'$ to $3'$ direction). Primers are short lengths of nucleotides, oligonucleotides, ranging from 8 to 24 basepairs in length for PCR depending on the desired specificity, which increases with primer length. PCR amplifies a region in the genome that is delimited by the oligonucleotide primers chosen for the initiation of DNA replication (Figure 3). The sequence of steps in the replication of genetic material is controlled by changes to the temperature of the solutions. The temperature of the solution is first raised (usually 94–95°C) to denature the two strands of DNA. Then it is lowered to allow primers to anneal to the template DNA (typical range is 50–65°C). It is then raised to a temperature optimal for activity of the thermal tolerant polymerase (e.g., Taq polymerase is best from 75 to 80°C) to synthesize new strands of DNA. The cycling of temperatures is repeated many times, resulting in amplification of discrete regions of DNA. This technology is used for nucleotide sequencing, synthesis of probes for hybridization experiments, detection of mutations, production of site-specific mutations, and synthesis of genes. As with all techniques, PCR-based methods also have limitations. The primers chosen for PCR can bias the populations amplified. The specificity of the primers is dependent on the sequence information available, thus different results can occur when different primer sets are used. Other limitations to the technique include: overestimates of diversity due to the formation of heteroduplexes, chimeric molecules, or sequence heterogeneities; and underestimation of diversity due to unequal amplification, the resolving power of the gel, or sequences comigrating in the gel. Nevertheless, this approach is very effective if the limitations of the primers are understood and PCR conditions are optimized to minimize errors in the PCR process. All the methods require access to a thermal cycler for PCR and gel electrophoresis units.

Community Richness

The identity of species and a crude measure of richness in a soil community can be determined by nucleotide sequence determination of rRNA genes. The method is independent of cultivation by using PCR directly to amplify rRNA genes from soil DNA extracts. The PCR products are cloned, unique clones identified by restriction fragment length polymorphism (RFLP), and then the nucleotide sequences determined. RFLP is sometimes referred to as amplified ribosomal DNA restriction analysis (ARDRA). The number of unique clones obtained can give an estimation of species richness in systems that are not too complex. The method is labor-intensive therefore an

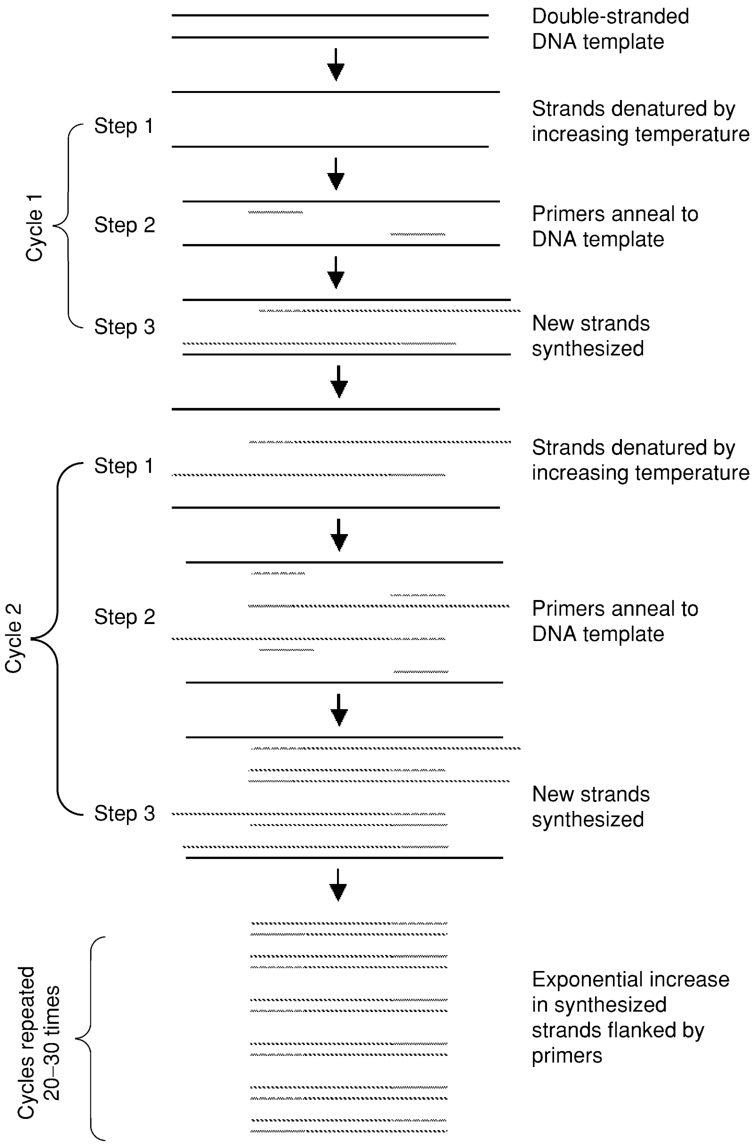


Figure 3 Steps in the polymerase chain reaction.

absolute number of species present in most soils is too time-consuming to obtain; for example, a variety of soils have been examined in which over 100 clones were screened and each clone was unique. However, the use of this approach has added new sequences to the rRNA database, revealed the presence of previously undescribed groups important in soil functions, and been used to determine unique oligonucleotide sequences to track specific strains in the community. For community analysis this method is best used in conjunction with other methods that can reduce the complexity of the sample. This method requires access to nucleotide-sequencing equipment or funds to send samples to a facility. Both are becoming more available to researchers.

PCR-Based Community Structure Analysis

To study differences in soil microbial community structure, researchers are using genetic profiles or fingerprints based on differences in rDNA sequences. The rRNA gene from the community mixture is first amplified by PCR. The primers chosen for PCR can be used to determine the level of discrimination between groups in the community. For example, primers can be used to amplify all Bacteria, or only the Actinomycetes. The products are then distinguished using gel electrophoresis, a means of separating nucleic acids by applying an electric current. The negative charge of the nucleic acids at neutral pHs allows the molecule to travel contained within a solid matrix, agarose or polyacrylamide gels, through an electric field. The nucleic acids typically separate according to their size, but denaturants can be added that will separate the molecules by their melting nature. The fractionated nucleic acids can be viewed after staining with a dye, e.g., ethidium bromide or SYBR green, and visualized by UV illumination. There are two electrophoresis methods that are being used to distinguish between PCR products of different populations present in a community. One approach, terminal restriction fragment length polymorphism (TRFLP), separates products by size and the other approach, by sequence composition, denaturing gradient gel electrophoresis (DGGE), or thermal gradient gel electrophoresis (TGGE).

Terminal restriction fragment length polymorphism TRFLP analysis is based on size differences of the terminal fragment of the rRNA gene after restriction enzyme digestion. In this method one or both of the primers used for PCR are labeled with a fluorescent molecule or a radioisotope. If both primers are labeled, then different fluorescence colours are used to distinguish the two ends. The PCR products are digested using a frequently cutting enzyme, e.g., *Rsa*I, *Hha*I. The more enzymes that are used to digest the fragments the finer a distinction can be made between numbers of different populations present. However, as the number of enzymes are increased, the time for analysis and interpretation increases. The products are separated in a polyacrylamide gel that can distinguish 1-bp differences in fragment length. The fluorescent fragments can then be visualized and counted. To facilitate the handling of a large number of samples and accuracy of fragment length distinction, automated nucleotide sequencing apparatuses are being used for analysis. Each fragment length represents a group that may consist of one or more populations. The band intensity can be used to quantify the number of PCR products represented by each fragment length, but this does not equate to the numbers of cells, because rDNA copy numbers vary between species. As more rRNA sequence information is accumulated, populations can potentially be identified based on their terminal fragment lengths.

Gradient gel electrophoresis This method uses electrophoresis in a denaturing gradient gel to distinguish PCR products of the same size by differences in their nucleotide sequence composition. There are two techniques that are based on the same principles, but one uses chemical denaturants, DGGE, while the other uses temperature, TGGE. The hydrogen bonds holding the two DNA strands are disrupted under denaturing conditions; the extent of disruption is dependent on the denaturant concentration and nucleotide sequence of the fragment. The extent of strand separation will affect the electrophoretic mobility of the DNA molecules. The different distances that the PCR products migrate can produce a distinct fingerprint profile. The band intensity can be used to quantify the number of PCR products. However, as stated earlier, potential PCR biases and variable rDNA copies should be considered when interpreting the data. In this method it is possible to identify populations represented by each band by nucleotide sequence determination of excised bands or by hybridization with specific probes.

Genetic fingerprinting methods are a useful first step in community analysis because multiple samples can be simultaneously and quickly analyzed to get an indication of richness, then subsequent analytical methods can be chosen if needed. They are an effective technique to discriminate between communities and monitoring dynamics of communities after environmental perturbation. Many soil communities (e.g., agricultural soils) have numerous populations in relatively equivalent proportions, thus resulting in a smear of bands that makes it difficult to differentiate communities (Figure 4, lanes 7 and 8). They can be differentiated if dominant bands are seen because of strong selective conditions, e.g., polyaromatic hydrocarbons (PAH), that give certain populations a competitive advantage or the overall richness has been decreased (Figure 4, lanes 1–6). Then the number of dominant bands in profiles with low complexity can be counted roughly to estimate richness. Care must be taken, because one band may represent one or more populations. Changes in the microbial community structure can be monitored by the appearance and disappearance of bands in the profiles over time. The presence and absence of bands can be used in quantitative analysis to estimate the similarity between two communities. However, care should be taken when interpreting these data because of the biases of PCR discussed earlier.

Data Analysis

The various methods described here need statistically valid analysis of the data produced. Similarity indices, such as Dice or Jaccard, are currently being used to quantify the relationship between two profiles. These data can be represented by cluster analysis, e.g., unweighted pair group method with arithmetic means (UPGMA), which provides a graphic illustration of the relationship between communities. Principal components analysis (PCA) is often used on large datasets. This method reduces the number of variables, determines whether there is a structure to the relationship between these variables, and can identify

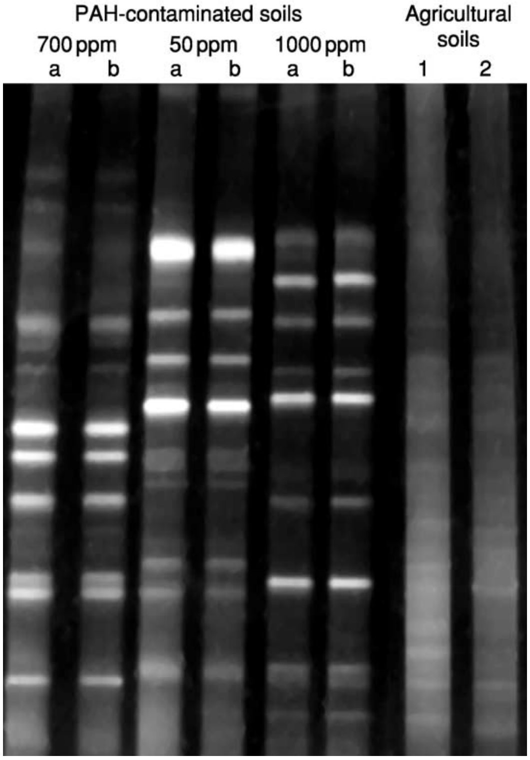


Figure 4 Soil community fingerprint profile of 16*S* rRNA gene V3 region amplified by polymerase chain reaction (PCR) and separated by denaturing gradient gel electrophoresis. Comparison of community profiles from soils contaminated with different concentrations of polyaromatic hydrocarbons (PAHs) to two different agricultural soils (labeled 1 and 2). Lanes a and b are replicate PCR reactions amplified from the same soil DNA extract. Primers used for PCR are bacterial primers BA338F (5′ AC TCC TAC GGG AGG CAG CAG 3′), with a GC clamp attached to the 5′ end (5′ CGC CCG CCG CGC GCG GCG GGC GGG GCG GGG GCA CGG GGG G 3′), and BA518R (5′ ATTACCGCGGCTGCTGG 3′).

main factors contributing to the variation. These techniques just begin to address specific research questions quantitatively, and now research groups are in the process of using these data to construct predictive statistical models of soil microbial communities. It is important that the proper experimental design and appropriate statistical analysis are employed to answer the research question being addressed.

Further Reading

Akkermans ADL, van Elsas JD, and de Bruijn FJ (eds) (1996) *Molecular Microbial Ecology Manual* (including supplement 5). Dordrecht, The Netherlands: Kluwer Academic Publishers.

El Fantroussi S, Urakawa H, Bernhard AE, Kelly JJ, Noble PA, Smidt H, Yershov GM, and Stahl DA (2003) Direct profiling of environmental microbial populations by thermal dissociation analysis of native rRNAs hybridized to oligonucleotide microarrays. *Applied and Environmental Microbiology* 69: 2377–2382.

Gerhardt P (ed.) (1994) *Methods for General and Molecular Bacteriology*. Washington, DC: ASM Press.

Hurst CJ, Crawford RL, Knudsen GR, McInerney MJ, and Stetzenbach LD (eds) (2001) *Manual of Environmental Microbiology*, 2nd edn. Washington, DC: ASM Press.

Johnson AR and Wichern DW (1998) *Applied Multivariate Statistical Analysis*, 5th edn. Upper Saddle River, NJ: Prentice Hall.

Kozdroj J and van Elsas JD (2001) Structural diversity of microorganisms in chemically perturbed soil assessed by molecular and cytochemical approaches. *Journal of Microbiological Methods* 43: 197–212.

Lederberg J (ed.) (2000) *Encyclopedia of Microbiology*, 2nd edn. London, UK: Academic Press.

Maier RM, Pepper IL, and Gerba CP (2000) *Environmental Microbiology*. London, UK: Academic Press.

Theron J and Cloete TE (2000) Molecular techniques for determining microbial diversity and community structure in natural environments. *Critical Reviews in Microbiology* 26: 37–57.

Tiedje JM, Asuming-Brempong S, Nusslein K, Marsh TL, and Flynn SJ (1999) Opening the black box of soil microbial diversity. *Applied Soil Ecology* 13: 109–122.

Torsvik V, Daae FL, Sandaa RA, and Øvreås L (1998) Novel techniques for analysing microbial diversity in natural and perturbed environments. *Journal of Biotechnology* 64: 53–62.

Kinetics

N S Panikov, Stevens Institute of Technology, Hoboken, NJ, USA

Introduction

Kinetics (Greek κινηματική, forcing to move) is a branch of natural science that deals with the rates and mechanisms of any process – physical, chemical or biological. Kinetic studies in soil microbiology cover all the quantitative and dynamic aspects of microbial life in the soil: growth and multiplication, mutation, survival and death of soil microorganisms, microbial interactions with the soil environment, other microbes, plants, and animals. Often studies of microbiological kinetics are part of interdisciplinary environmental research on soil processes on which microorganisms play a unique or essential role: organic matter decomposition, biogeochemical cycles of methane, S, N, Fe, P. Kinetics provides a theoretical framework for a better understanding of soil function and for the optimal design of environmental biotechnologies. These technologies are based on the release of beneficial microbes into the environment or activation of the indigenous microbial community and may benefit research into soil bioremediation, sustainable agricultural systems, and an integrated plant defense system. Recently the kinetics of microbial processes has increasingly focused on the construction of various 'soil modules' in large-scale ecosystem and biosphere models aimed at quantitative prediction (e.g., global climatic changes) and optimal resource management.

Contrary to simple monitoring of soil dynamics (time-series record of the soil state), kinetic studies require an understanding of the mechanisms underlying processes of interest. For example, a complex process such as plant litter decomposition must be viewed as the interplay of several simpler reactions: the conversion of individual litter components, growth of particular degraders, and specific enzymatic transformations. Ideally, mechanistic studies infer the coupling of experimental measurements with simulation mathematical models. These models formalize postulated mechanisms, so that the comparison of observations and the model's predictions allows incorrect hypotheses to be discarded.

Microbial kinetics is tightly linked to growth stoichiometry (Greek στοιχείο, element), which is the quantitative relationship between substrates and the products of microbial growth. In practical terms, stoichiometry mainly addresses problems of a static nature, like 'how much?' and 'in what proportion?' while kinetics considers the dynamics and deals with questions like 'how fast?' and 'by which mechanism?'

The Main State Variables and Growth Parameters

There are two types of kinetic models, deterministic and stochastic. The former describe clear, determined, and regular processes. The latter models deal with random processes. The main variables used in deterministic models are the same as in chemical kinetics – concentrations of cell mass, substrates, and products – while stochastic models consider instead probabilities, frequency distributions, and variance. For example, a deterministic model can describe the process of N_2-fixation as a function of energy and nutrient supply to the rhizosphere, while a stochastic model considers the probability of rhizobia to infect legume roots for a given frequency distribution of bacteria across the soil space. Any real soil process has both deterministic and stochastic components, so ideally the kinetic model should be a 'hybrid' and contain a deterministic core with a stochastic interface, generating probability distributions and confidence corridors rather than single values or curves.

The major state variables of kinetic models describing microbial processes in the soil are summarized in Table 1 and described in detail below.

Amount of Soil Microorganisms

Amount of soil microorganisms can be expressed as biomass (x) or cell numbers (N) of unicellular organisms (bacteria, yeasts, protozoa, fungal spores) per unit soil mass, volume, or surface area. The filamentous organisms (fungi, actinomycetes) are characterized by length (L) of mycelium and number of hyphal tips, n, initiated by hyphal branching. Note that n is equivalent to N, as soon as branching is equivalent to cell division.

The relationship between x and N or L is not straightforward because the mass of individual cells and hypha varies among different organisms and even for the same species depending on growth conditions. Generally, a plentiful supply of nutrients fosters the formation of bigger cells or wider filaments while starvation causes the formation of smaller cells and thinner filaments. To estimate soil biomass from direct microscopic counting, one should generate several arrays of data: the number $\mathbf{N} = \{N_1, N_2, \ldots N_n\}$ of cells with specifying biovolume $\mathbf{V} = \{V_1, V_2, \ldots V_n\}$, where N_i is the number of cells in the ith volume class V_i and n is the number of such classes, the volumes being calculated from lengths and widths of all detected cell shapes (spherical, ellipsoid, rod-shaped, irregular). For filamentous microorganisms one should have arrays of hyphal length

Table 1 The major dynamic variables used in deterministic kinetic models

Variable	*Symbol*	*Dimension (examples)*
Density/concentration of cell biomass	X	μg cell mass (g soil^{-1}) or g cell mass m^{-2} or μg cell mass (cm^3 soil)$^{-1}$
Cell number[a]	N	10^6 cell (g soil)$^{-1}$ or 10^6 cell (ml soil solution)$^{-1}$
Single-cell mass	m	g dry weight per single cell
Mycelium length[b]	L	m (g soil)$^{-1}$
Tips number	n	10^6 (g soil)$^{-1}$
Concentration of limiting substrate	s	mg (g soil)$^{-1}$, g m^{-2} or g l^{-1} of soil solution
Concentration of product (CO_2, CH_4)	p	mg (g soil)$^{-1}$, g m^{-2} or g l^{-1} of soil solution

[a] For unicellular organisms (bacteria, yeasts).

[b] For filamentous organisms (fungi, actinomycetes).

$\boldsymbol{L} = \{L_1, L_2, \ldots L_m\}$ and corresponding hyphal diameters $\boldsymbol{D} = \{D_1, D_2, \ldots D_m\}$. The total soil biomass, x, can be calculated as a sum

$$x = \boldsymbol{V} \times \boldsymbol{N} \times (1 - \Theta) + 1/4\pi \boldsymbol{D}^2 \times \boldsymbol{L} \times (1 - \Theta) \qquad [1]$$

where Θ is the water content in wet cell mass (usually taken from lab data).

Advanced microscopic (ultraviolet, electron, and atomic force microscopy) and computer-aided image-analysis techniques make cell counting and sizing less tedious. Additionally, fluorescence *in situ* hybridization (FISH) allows targeting of specific taxonomic or functional groups based on the use of RNA-oligonucleotide probes. The total soil biomass, x, can be quantified by extracting from soil samples and quantifying one of the following integral and unique cell constituents: (1) DNA; (2) adenosine triphosphate (ATP); (3) microbial phospholipids; (4) microbial cell wall components (chitin for fungi and muramic acid for bacteria); or (5) the sum of labile/extractable C released after soil fumigation with CCl_3.

Selection of x or $N(L)$ depends on the specific research question. Biomass, x, has the obvious advantage in studies of carbon and nutrient cycling, while cell number, N, is preferred in population studies when, say, mutation or plasmid transfer is crucial for an understanding of the studied soil processes.

Substrate Concentration

Substrate concentration in the soil, s, stands for the amount of some essential soil constituents used by soil microbes for growth and maintenance. There are two groups of such constituents: (1) catabolic substrates, which are sources of energy; and (2) anabolic or conserved substrates, which are sources of biogenic elements forming cellular material. Examples of catabolic substrates are H_2 for lithotrophic hydrogen bacteria, NH_4^+ and NO_2^- for nitrifying bacteria, sulfur for sulfur-oxidizing bacteria, and oxidizable or fermentable organic substances for heterotrophic species. Their consumption is accompanied by oxidation and dissipation of chemical substances into waste products which are no longer reusable as an energy source (H_2O, NO_3^-, SO_4^{2-}, CO_2, etc.). (Fermentation products like acetate, ethanol, butyrate, H_2 etc. seem to be an exception as they do contain reusable oxidation potential, but it is not available under anaerobic conditions supervising fermentation.) After uptake, anabolic substrates are incorporated into *de novo* synthesized cell components, being conserved in biomass; thus, sometimes anabolic substrates are referred to as conserved substrates. In contrast to catabolic substrates, the conserved substrates can be reabsorbed after excretion or cell lysis. They include nearly all noncarbon sources of biogenic elements (N, P, K, Mg, Fe, and trace elements), CO_2 for autotrophs, as well as amino acids and growth factors.

Two variables, cell mass x and substrate concentration s, are linked via stoichiometric parameter growth yield:

$$Y = -\frac{dx}{ds} \approx -\frac{\Delta x}{\Delta s} \qquad [2]$$

where Δx is the increase in microbial biomass consequent on utilization of the amount Δs of substrate. Dividing both parts of eqn [2] by $x\delta t$, gives the relationship between growth rate and substrate production:

$$Y = -\frac{\delta x}{\delta s} = -\frac{\delta x}{x\delta \mathrm{t}} : \frac{\delta s}{x\delta \mathrm{t}} = -\mu/q \qquad [2a]$$

where μ is specific growth rate and q is specific rate of substrate consumption.

The reason for Y variation is different for catabolic and anabolic substrates. In the case of energy sources, some fraction of the total substrate flux is diverted from growth *per se* to meet so-called maintenance functions, such as resynthesis of self-degrading cell proteins, nucleic acids and other macromolecules, osmotic work to keep the concentration gradient between cell interior and environment, and cell motility. The energy balance is described by the following equation:

$$\underset{q}{\text{Total energy source consumed}} = \underset{\mu/Y^{\max}}{\text{Consumption for growth}} + \underset{m}{\text{Consumption for maintenance}} \qquad [3]$$

where m is the maintenance coefficient or the specific rate of catabolic substrate consumption by nongrowing cells (i.e., $m = q$ when $\mu = 0$). (With some rare exceptions (fungal exospores and bacterial cysts), the microbial cells are not stable at $\mu = 0$ and either grow ($\mu > 0$) or decay ($\mu < 0$). Therefore the maintenance coefficient is found by linear extrapolation of a series of $q(\mu)$-measurements to the point where μ is zero.) Under chronic starvation the maintenance coefficient, m, becomes progressively smaller in contrast to higher values associated with intensive growth; as a result, when $\mu \to 0$, the growth yield Y approaches a low, nonzero value $Y^{\min}$ (**Figure 1**).

There is also wasteful oxidation of substrate under at least three specific circumstances: (1) when growth is nutrient-limited and energy-sufficient; (2) when starving cells are brought to rich nutrient medium (famine-to-feast transition); and (3) when cells are exposed to uncoupling inhibitors. In all these cases,

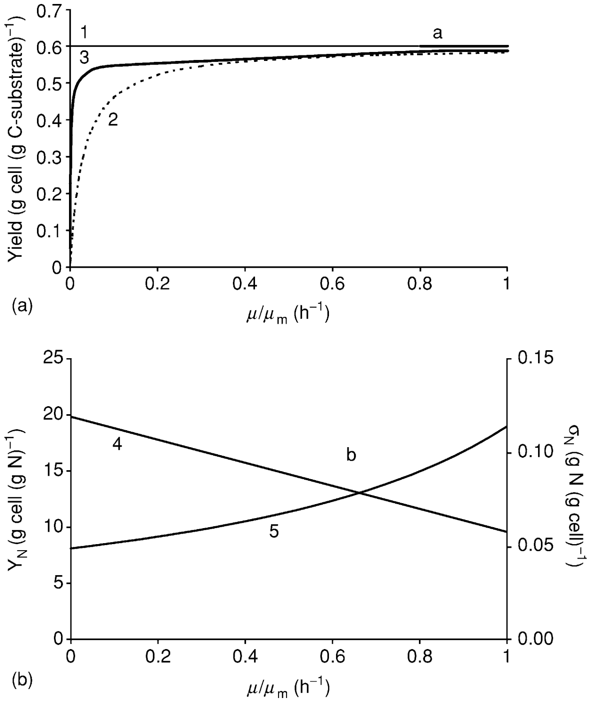


Figure 1 Variation of the cell mass yield on catabolic (a) and anabolic (b) substrates related to specific growth rate of soil microorganisms. Curves calculated assuming no maintenance (1), constant maintenance coefficient (2), and adaptive decrease in maintenance in slowly growing cells (3), Curve 4 is cell yield per g N, and curve 5 is cell quota, g N/g cell. Curves 3–5 were calculated using SCM.

the cell's catabolic machinery produces more energy that can be used for ATP generation. Such wasteful catabolism should be differentiated from maintenance *per se*.

Cell yield on anabolic substrates varies mainly as a result of alterations in biomass chemical composition as expressed by the parameter σ_s, the intracellular content of deficient element or cell quota. The variation in N content in bacteria from 5 to 15% gives the σ_s diapason 0.05–0.15 g N (g cell mass)$^{-1}$. For most known cases, the quota σ_s increases in parallel to growth acceleration, because the higher growth rate requires higher intracellular content of proteins and RNA (containing N, P, S) as well as K^+, Mg^{2+}, and vitamins participating in all primary metabolic reactions. The yield and cell quota are inversely related to each other, e.g., low N-content ($\sigma_N = 0.05\,\mathrm{g\,N\,g^{-1}}$) corresponds to high cell yield ($Y_N = 1/\sigma_N = 20$ g cell per g of N utilized), high N-content in rapidly growing cells can be attained only with low cell yield ($Y_N = 1/0.15 = 6.67$ g cell per g N) (Figure 1).

Products

There may be intermediate and end metabolic products accompanying microbial growth. The typical intermediates are organic acids forming in glycolysis/Etner–Doudoroff pathways or various fermentations:

1. glucose → 2 pyruvate$^-$ + 2H
2. glucose → gluconic acid → pyruvate
3. pyruvate → acetate

The only end product of aerobic decomposition is CO_2, whereas products formed under anaerobic conditions include organic acids, alcohols and ketones, CO_2, CH_4, and various reduced inorganic compounds (S^{2-}, Fe^{2+}). The CO_2 evolution rate is probably the most popular parameter to characterize soil biological activity due to the wide availability of gas analyzers (gas chromatography, infrared, and mass spectrometry). In particular cases of aerobic decomposition, the measured CO_2 evolution is a significant fraction of the C budget, as described by the following equation:

$$\overset{\text{Total carbon consumed}}{\Delta C - s} = \overset{\text{C incorporated into cell}}{\Delta C - x} + \overset{\text{C oxidized to CO}_2\text{ to provide energy}}{\Delta C - CO_2} + \overset{\text{C in extracellular products}}{\Delta C - p} \qquad [4]$$

Dividing by time increment δ/t gives:

$$\overset{\text{Uptake of C-substrate}}{\frac{\delta s}{\delta t}} = \overset{\text{Microbial growth}}{\frac{\delta x}{\delta t}} + \overset{\text{Respiration}}{\frac{\delta[CO_2]}{\delta t}} + \overset{\text{Products formation}}{\frac{\delta p}{\delta t}} \qquad [4a]$$

Thus, the microbial consumption of C-substrate is the sum of growth, respiration, and product formation. Based on such C-balance, we can reconstruct the missed variables. For example, cell mass x in a laboratory soil incubation study can be calculated as the difference between the amount of initially added substrate and current s, p, and CO_2 values. (The priming effect of added organic matter on decomposition of soil humus must be evaluated independently – it is usually small enough. The priming is fully excluded when using labeled amendment.) However, it would be incorrect to identify the CO_2 evolution rate alone with total soil decomposition, as some degraded C can be (and really is) incorporated into soil organisms or their extracellular products.

The Models of Microbial Growth

Exponential Growth

Under favorable growth conditions (temporary excess of nutrient substrates, absence of inhibition), the bacterial growth rate should be proportional to the instant cell mass, x, the quotient μ remaining constant:

$$\frac{\delta x}{\delta t} = \mu x \qquad [5]$$

The integration of eqn [5] at initial condition $x = x_0$ at time $t = 0$, gives the exponential equation:

$$x = x_0 e^{\mu t} \quad \text{or} \quad \ln x = \ln x_0 + \mu t \qquad [5a]$$

The rate of CO_2 or CH_4 production (gas production rate: GPR) is proportional to the growth rate of the respective group of soil microorganisms:

$$\overset{\text{GPR}}{\frac{\delta p}{\delta t}} = \overset{\text{Yield factor}}{\frac{1}{Y}} \times \overset{\text{Cell growth}}{\frac{\delta x}{\delta t}} = \frac{\mu x}{Y} = \frac{\mu x_0}{Y} e^{\mu t} = A e^{\mu t} \qquad [6]$$

where p is gaseous product concentration, expressed for simplicity in the same units as biomass, say mg C per g soil. Integration of eqn [3] under initial condition $p = 0$ at $t = 0$ gives:

$$p = \frac{x_0}{Y}\left(e^{\mu t} - 1\right) = B \times \left(e^{\mu t} - 1\right) \qquad [6a]$$

Comparing eqns [5a] and [6a], we see that x and p dynamics are not identical, although both imply exponential cell growth; the $p(t)$ curve deviates from the exponential law, especially at small t. In particular the '$\ln x - t$' plot is linear, but '$\ln p - t$' is not. As t increases, the difference becomes smaller and smaller.

In contrast to the exponential growth equation described above, the logistic equation sets an upper limit for microbial growth by empirically assuming that the specific growth rate μ is dependent on microbial concentration:

$$\frac{\delta x}{\delta t} = \mu(x)x = \mu x\left(1 - \frac{x}{x_m}\right) \qquad [7]$$

where x_m is the upper limit of cell mass. In ecological and demographic literature, x_m is called the carrying capacity of ecosystem, K. This equation produces the S-shaped growth curve frequently observed in nature. However, this equation is no more than empiric, and parameters μ and x_m are not 'fundamental,' i.e., they cannot be measured and tabulated for particular soils or ecosystems or predicted before actual observation.

Monod Model

This model, developed by the famous French scientist in 1941, is still very popular due to its elegant simplicity. There are two basic premises: (1) the specific growth rate, μ, is dependent on the concentration of one nutrient component which happens to be the most deficient, the so-called limiting substrate: and (2) biomass formation is linked to substrate uptake by mass-conservation conditions (eqn [2a]). Therefore, according to these premises:

$$\begin{aligned} \frac{\delta x}{\delta t} &= \mu(s)x = \mu_m \frac{s}{K_s + s} x \\ \frac{\delta s}{\delta t} &= -\frac{1}{Y}\frac{\delta x}{\delta t} \end{aligned} \qquad [8]$$

Eqn [8] contains three parameters: yield Y, maximal specific growth rate μ_m, and saturation constant K_s (substrate concentration at which $\mu = 0.5\ \mu_m$), which can be thought of as the 'ID' for particular organisms and used to predict their growth dynamics.

Remarkably, this model was used to develop a chemostat theory before actual experiments with continuous culture had been undertaken, one of the very rare events in the history of mathematical biology! The model predicted a number of counterintuitive features of the chemostat, e.g., that specific growth rate μ can be set up by an experimentalist by changing the medium flow at any values between 0 and μ_m (previously, exponential growth was believed to occur only at $\mu = \mu_m$) and that μ-values do not depend on the feed-substrate concentration and are governed solely by the residual substrate concentration in the culture.

However, the Monod model fails to explain a number of essential growth phenomena observed experimentally, namely: lag-phase, death of starving cells, product formation and any kind of adaptive changes in microbial population, such as induction-repression of enzymes, yield variation, and changes in the cell RNA content. These gaps were filled in by so-called structured models.

Structured Models

Structured models explicitly describe variations in cell composition. They usually include mass balance equations not only for external substrate(s) but also several intracellular components, $C_1, C_2, \ldots C_n$. For each variable C_i a differential equation is written which takes into account all sources, r_+, and sinks, r_-, as well as its dilution due to cell mass expansion (growth):

$$\frac{\delta C_i}{\delta t} = r_+(s, C_1 \ldots C_n) - r_-(s, C_1 \ldots C_n) - \mu C_i \qquad [9]$$

The earliest structured models accounted for no more than 3–5 cell constituents, e.g., total cell proteins, RNA and DNA, reserved polysaccharides, and ATP-pool. Modern models contain up to hundreds and even thousands of internal variables taken directly from available genomic/metabolic databases. Currently the challenge is to create a virtual cell (so-called cell *in silico*) that is complex enough to function as a real cell. However the predictive capability of these intricate models is rather modest: they are still a 'caricature parody' of the real cell, but too complex to be efficiently studied mathematically (stability analysis, parameters identification) or to improve our understanding of the biosystem. The best choice of a mathematical model lies, apparently, midway between the unstructured and highly structured models outlined here. One of the known examples is the synthetic chemostat model (SCM).

Synthetic Chemostat Model

According to the SCM, microbial growth occurs by converting exosubstrate, S, into a number of cell macromolecules, X', via a pool of intermediates, L, part of which are respired to CO_2 (Figure 2).

Macromolecular cell components are susceptible to degradation (turnover), and intermediates L can leak out. The array X', the cell composition, is not fixed and varies in response to the changing environment. The heart of the SCM is the solution of the problem: how to characterize these variations without going too deeply into details?

For this purpose, all the macromolecular cell constituents are divided into two groups: Primary cell constituents necessary for intensive growth (P-components), and components needed for cell survival under any kinds of growth restriction (U-components). The content of P-components (ribosomes and all enzymes of the primary metabolic pathways) increases in parallel with growth acceleration. The contribution of U-components under good growth conditions decreases (to comply with conservation conditions $\mathrm{P}+\mathrm{U}=$ constant), and attains the maximum under chronic environmental stress to improve cell resistance. Typical U-components are enzymes of secondary metabolism, protective pigments, reserved substances, and transport systems of high affinity. An interdependent variation of individual P- and U-components is approximated by linear function of some master variable r:

$$\mathrm{P}+\mathrm{U}=\begin{vmatrix} P_1 \\ P_2 \\ \dots \\ P_\mathrm{n} \end{vmatrix}+\begin{vmatrix} U_1 \\ U_2 \\ \dots \\ U_\mathrm{m} \end{vmatrix}=\begin{vmatrix} P_1^{\min} \\ P_2^{\min} \\ \dots \\ P_\mathrm{n}^{\min} \end{vmatrix}+\begin{vmatrix} U_1^{\min} \\ U_2^{\min} \\ \dots \\ U_\mathrm{m}^{\min} \end{vmatrix} + r\begin{vmatrix} P_1^{\max}-P_1^{\min} \\ P_2^{\max}-P_2^{\min} \\ \dots \\ P_\mathrm{n}^{\max}-P_\mathrm{n}^{\min} \end{vmatrix}+(1-r)\begin{vmatrix} U_1^{\max}-U_1^{\min} \\ U_2^{\max}-U_2^{\min} \\ \dots \\ U_\mathrm{m}^{\max}-U_\mathrm{m}^{\min} \end{vmatrix} \quad [10]$$

where $P^{\max}$, $U^{\max}$, and $P^{\min}$, $U^{\min}$ are respectively upper and lower boundaries for P and U adaptive variations, and r is the scalar function, not array! The r-value depends directly on environmental factors, e.g., on the limiting substrate concentration and is included in all kinetic expressions (q_s, m, μ) to simulate the combined effects of current environmental factors and the cellular physiological state, as determined by past growth conditions.

The predictive and explanatory capabilities of SCM are higher than that of more complex structured models. It usually gives good agreement with most available experimental steady-state and transient growth data, such as batch culture from lag- to decline phase or continuous culture after abrupt changes in dilution rate, pH, or temperature. For ecological applications, SCM realistically describes and predicts not only growth *per se* but also many other dynamic phenomena: survival dynamics under starvation, formation of dwarf cells under growth restriction, the adaptive adjustment in cell maintenance requirements, variation of growth potential and affinity to substrate, and utilization of substrate mixture. Below we will use SCM to explain microbial growth kinetics both in culture and in the soil.

Growth Kinetics *In Situ* and *Ex Situ*

Soil as a microbial habitat is a system of relatively independent microsites, each of which can be drastically dissimilar in the amount and nature of growth substrates, aeration level, texture, and moisture. However, there are several general features which

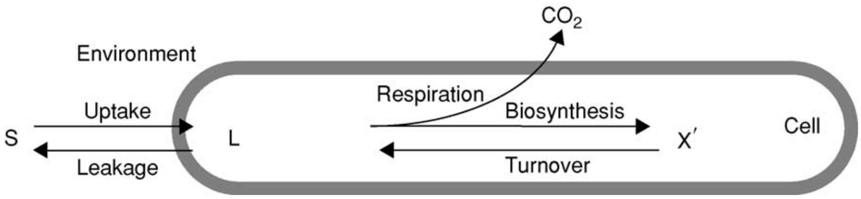


Figure 2 Flow diagram of major cell processes used to stimulate microbial growth by basic SCM.

Table 2 Matrix of growth patterns *in situ* and *ex situ*

	Substrate input			
	Continuous (1)		*Single-term (2)*	
Spatial organization	*Cell removed (α)*	*Cell retained (β)*	*Cell removed (α)*	*Cell retained (β)*
Homogeneous (a)	1aα	1aβ	2aα	2aβ
Heterogeneous (b)	1bα	1bβ	Forbidden combination	Forbidden combination

are most essential for determining the population dynamics of microorganisms occupying a respective microsite:

1. Substrate input may be continuous (1) or discontinuous (2).
2. Growing cells can be retained within a microhabitat (α) or eliminated (β). Examples of elimination include: cell washout, predation (consumption of microbial cells by protozoa, nematodes, microarthropods) or lysis by parasites, active migration due to taxis (motile bacteria) and tropism (vectorized apical growth of hyphal organisms).
3. Habitats vary in spatial organization, being (a) homogeneous and (b) heterogeneously organized. In the first type of system, there is an even or at least random distribution of cells, substrate, and metabolic products in space, whereas in the second type of habitat there are gradients in x, s, and p. The homogenization in soil occurs due to tillage and fossorial animals. Homogeneous microenvironments are also formed when the growth substrates are mobile: gases and volatiles.

These three independent characteristics of ecotopic conditions produce eight possible combinations (Table 2): two of them are logically forbidden because any kind of spatial heterogeneity implies continuous rather than momentary single-term delivery of substrate from some parts of habitat across the *s* concentration gradient. Figure 3 depicts several characteristic microbial habitats occurring in the soil. The most spacious and poorest/oligotrophic habitat is called the dispersion zone (1aβ); the subsoil and patches of bare topsoil are devoid of plant roots or fresh litter. Numerous microbial populations inhabiting these habitats grow very slowly on volatile or readily soluble compounds, which continuously diffuse from other soil loci where monomeric concentration is high due to intensive decomposition (plant litter) or excretory activity of plants (rhizosphere). The lack of elimination (no motility in the majority of soil bacteria and no predation due to low prey density) combined with slow but uninterrupted continuous growth eventually results in a significant buildup of a half-dormant cell mass.

The highest and most continuous microbial activity is localized in the soil litter layer, rhizosphere, and digestive tract of soil animals (habitats of the types 1aβ–1bβ). The C-substrates for microbial populations are the monomeric labile compounds (sugars and organic acids) derived from plants either as root exudation in the rhizosphere or released by extracellular hydrolytic enzymes from the lignocellulose and other polymeric material in plant litter. The litter layer on the soil surface is formed from deposition of aboveground plant remnants, while belowground plant senescence (root litter and root sloughing) provide microbial C-substrates in the rhizosphere. Root exudation is closely related to plant photosynthesis and changes diurnally, whereas hydrolytic release is monotonous. Spatially, all these habitats are rather heterogeneous; however, the random distribution of microloci combined with macroscopic sampling size allows us to use homogeneous kinetic models. Clear vertical gradients are formed in the litter layer from uncolonized fresh plant debris on the top to highly decomposed sublayer at the interface with mineral soil. At the microscale, the spatial heterogeneity of microbial colonies (mainly fungi) is manifested in its differentiation into growing extension zone and the nongrowing reproductive compartments (Figure 4). Within the rhizosphere, there are also several spatial gradients of different scales: (1) the vertical gradient of the root phytomass, which reflects the spatial pattern of belowground allocation of photosynthate; (2) the horizontal gradient between distant trees or tussocks; and (3) the micro-scale gradient around root hair with maximal concentration of microbial cells and substrates on the root surface (rhizoplane) and exponential decline outwards. (Some researchers found a statistically significant decline in cell density of rhizosphere bacteria on the plant surface, probably due to excretion of antibiotic compounds by plants.)

Soil millipedes, isopods, some earthworms and other primary decomposers inhabit the litter layer and feed on plant debris. The ingested lignocellulose material is mechanically disrupted by the mandible,

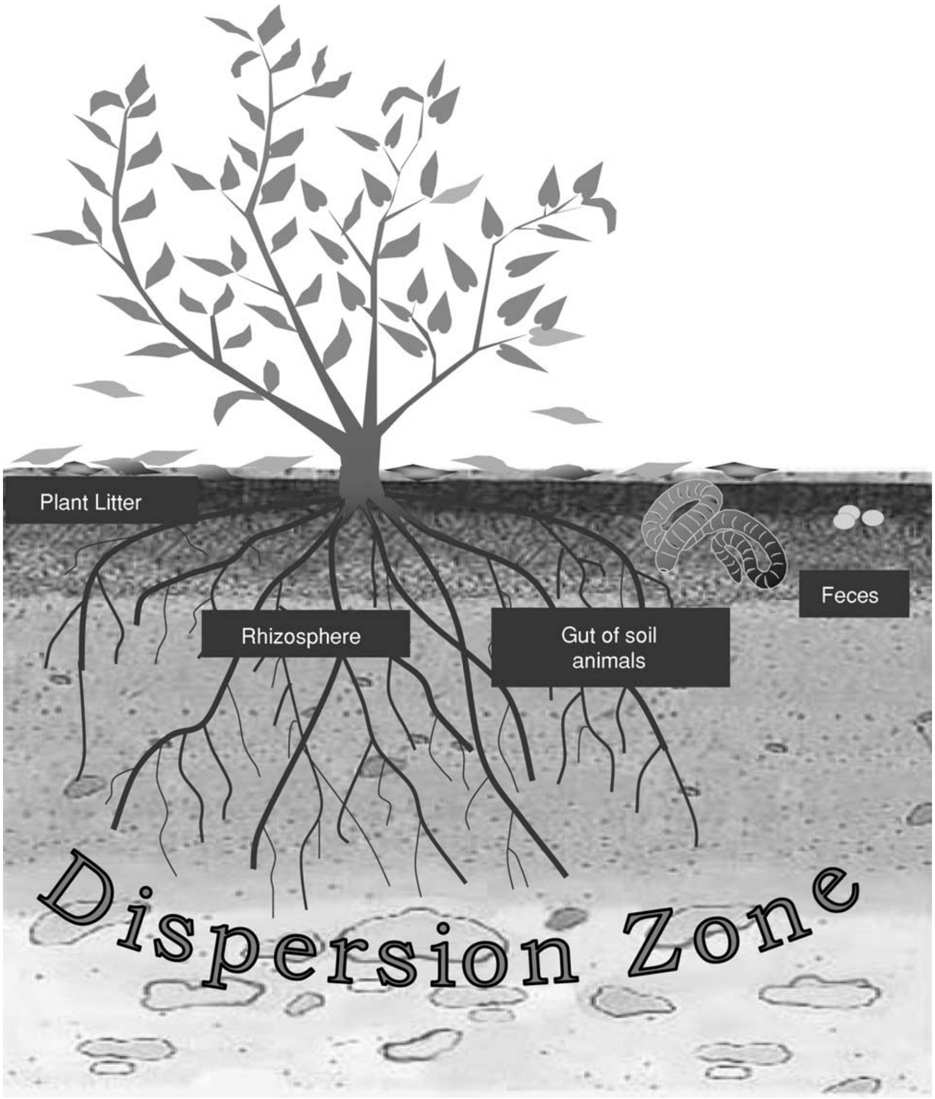


Figure 3 The principal microbial habitats in the soil (see text for explanation).

moistened with saliva, and then passed to the mid- and hindgut. The digestive tracts of various soil invertebrates harbor not only specific symbionts, but also the normal free-living microorganisms occurring in soil or plant litter, e.g., *Pseudomonas*, *Flavobacterium*, *Vibrio*, *Enterobacter*, and *Streptomyces*. The acceleration of their growth in the hindgut is due to favorable conditions: neutral pH, optimal moisture, elevated concentrations of nutrient and growth factors (amino acids, peptides, vitamins), as well as the continuous input of fresh substrate and concomitant removal of digestion products (glucose) to prevent the negative feedback (so-called catabolic repression) on cellulase synthesis. Due to peristaltic motion the content of the gut is mixed and homogeneous. Secondary decomposers (i.e., the earthworms *Allolobophora chlorotica*) feed on amorphous humus containing bacteria and fungal mycelia; they eliminate some microbial species and greatly stimulate growth of others.

Discontinuous, explosive microbial growth occurs within hot spots initiated by sudden increases in the availability of organic substrates/nutrients in the soil (2aα–2aβ), such as feces and carcasses of animals, rain washing of organic matter from plant foliage, dry–rewet or freeze–thaw cycles, manure application, and soil fumigation. Growth is usually accompanied by elimination in the form of grazing, myco- and bacteriolytic activity, as well as the active migration of microbial cells.

The dynamics of microbial growth in any type of habitat is described by the mass balance equations:

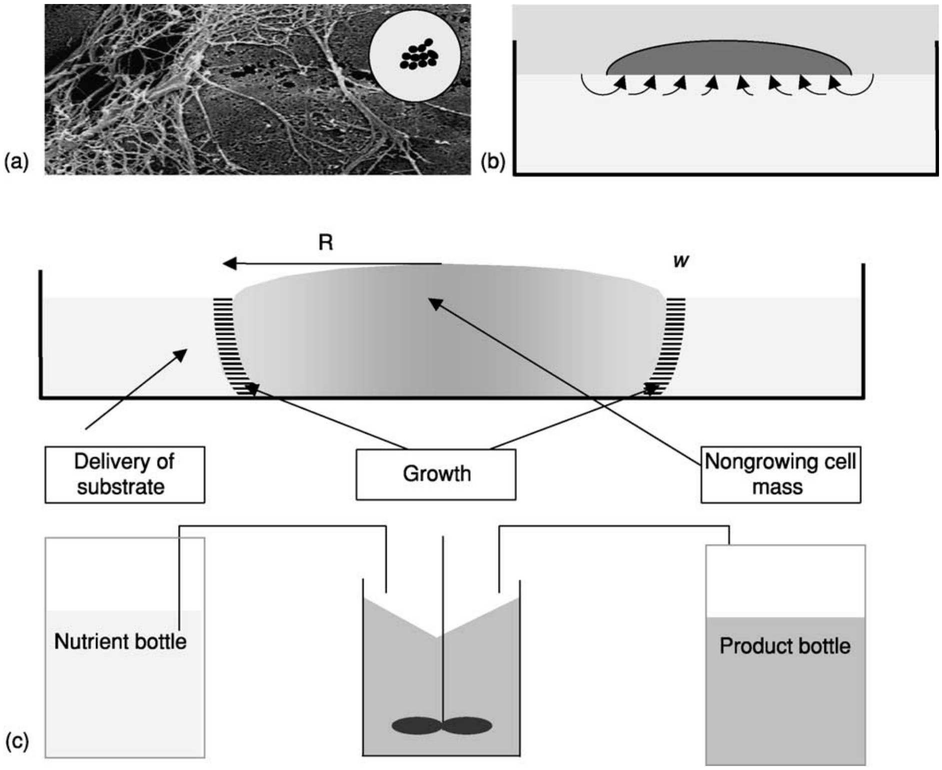


Figure 4 The colony growth: (a) in the soil; (b) bacterial growth on agar plate; (c) fungal growth in nutrient agar. Note that bacteria grow only on the agar surface and the colony expansion is controlled by diffusion of substrates from outside the colony. The fungi and actinomycetes are able to penetrate into the depth of the agar layer, so their mycelium expansion is not dependent on nutrient diffusion within agar layer. The fungal colony follows a chemostat-type growth pattern, being: (1) continuous; (2) steady-state; and (3) limited by substrate availability. The role of the fermentation vessel of the conventional chemostat is played by the peripheral zone of the colony; the product bottle is analogous to the central part of the colony, while the pump is substituted by the chemotropic movement of hyphae tips along the substrate concentration gradient.

$$\begin{array}{ccccccccc} \text{Net change} & & \text{Substrate input} & & \text{Substrate loss} & & \text{Uptake for growth} & & \text{Uptake for maintenance} \\ \frac{\delta s}{\delta t} & = & F(t) & - & G(s) & - & \frac{\mu(s)x}{Y} & - & mx \\ \text{Net change} & & \text{Immigration} & & \text{Elimination} & & \text{Growth} & & \\ \frac{\delta x}{\delta t} & = & V(t) & - & H(x) & + & \mu(s)x & & \end{array} \qquad [11]$$

To specify kinetic expressions, we can rely on kinetic models developed for the laboratory cultivation system which is closest to the respective soil habitat (Table 3). For instance, growth of soil bacteria in the intestine of earthworms is remarkably similar to a tubular (plug-flow) reactor, the rhizosphere community is related to a chemostat or phased-batch culture with injection of the growth-limiting substrate once per day (simulation of photosynthesis-dependent root exudation), the slow microbial growth in dispersion zone is adequately simulated by dialysis culture, and growth of fungi degrading plant litter contains kinetic patterns similar to colonies developing on agar.

Growth models based on SCM for various homogeneous cultivation systems are shown in Figure 5, and colonial growth is illustrated in Figure 4. One of the best criteria of adequacy of the selected cultivation system is specific growth rate, μ, of the studied microbial population. If the values of μ for *ex situ* and *in situ* populations coincide, then the cultivation technique may be regarded as appropriate.

Diversity of Growth Characteristics: Life Strategy Concept

In soils, there are always multiple microbial populations performing the same function. They have different growth parameters and regulators, but they require and compete for the same nutrient substrates. Their coexistence, which contradicts the well-known principle of competitive exclusion, can be explained by at least two factors:

1. spatial segregation of soil microsites, which leads to a lack of direct interactions between potentially competing species in the soil

Table 3 Examples of typical soil habitats and their counterparts in laboratory cultivation systems

		Soil habitats		
Kinetic type	*Lab cultivation system*	*Locus*	*Substrate flux*	*Elimination factor*
1aα	Chemostat, turbidostat, pH-auxostat	Rhizosphere	Rhizodeposition: exudation and sloughing	Root extension, predation by protozoa and nematodes
1aβ	Fed batch culture, dialysis culture	Dispersion zone	Diffusion of gases, volatile and soluble organic matter	No or very weak
1bα	Plug-flow, colonies on plate	Digestion tract of invertebrates	Predigested plant litter	Excretion
1bα	Colonies, solid-phase fermentation	Plant litter	Litter fall, root sloughing	Predation by microarthropods, apical hyphal extension
1bβ	Column packed with microbe attached, biofilms	Vertical network of the soil water channels	Trickling solutions	Cell motility
2aα	Phased culture	Temporary hot spots of microbial activity	Excrements and corpses soil animals, manure	Predation/grazing and exposure to harmful factors
2aβ	Simple batch	Hot spots without intensive microbial grazing	Washout of labile organic matter with rainwater	No or weak predation[a]

[a] Low probability of intense predation if no buildup of microbial biomass (prey population density) occurs.

2. perpetual environmental changes which cause variation in the direction/intensity of natural selection, so the fittest competitor today can be different from the day before, none of them being excluded completely

Studying microbial communities from a kinetic perspective can describe quantitatively why the pressure of natural selection can differ. One can distinguish the following three types of natural selection which correspond to three types of life strategy:

1. *K*-selection for high competitive power ('lions'); operates in climax ecosystems
2. *r*-selection for reactivity and ability to recover after perturbations ('jackal'); operates on the pioneer stages of succession
3. *L*-selection for resistance to stress-factors ('camel'); operates under adverse environmental conditions (Figure 6)

Figure 7 summarizes results of kinetic studies of the typical microbial *r*-, *K*-, and *L*-strategists. The *Enterobacteria*, *Pseudomonas*, Baker and Fodder yeast are mainly *r*-selected organisms. Their dynamic behavior is erroneously considered to be typical for all microbes: rapid and balanced growth, short lag-periods, and smooth transitory processes. They dominate in those soil habitats that are frequently 'rejuvenated' to the pioneer succession stage: hot spots of substrate amendment, animal guts and feces, rhizosphere with diurnal fluctuations in exudation rates and perpetual changes in 'address' of exudation loci due to apical extension of root hair. Plating is a good simulation of such hot spots; that is why *r*-selected species are easily isolated from soil.

K-selected bacterial species are much less amenable to isolation and cultivation. Probably most unculturable microbial species belong to this type of life strategy. When cultivated under artificial laboratory conditions, they are fastidious and unpredictable (a 'lion' in captivity). The best option for their cultivation is continuous culture with cell retention: fed-batch, dialysis culture, or batch culture with C-substrate delivered via gas phase (volatile C-substrates, such as ethanol or volatile fatty acid). Under these cultivation conditions the *K*-selected species display high yield and almost 100% viability, even at extremely slow cell division (generation time up to months). In other cultivation systems, their growth patterns are characterized by: (1) biphasic growth on the single substrate in batch culture (false diaxia); (2) long survival under starvation; and (3) oscillations and bistability (two or more steady states at one dilution rate) in chemostat. To simulate mathematically the described abnormal behavior of *Arthrobacter* and other oligotrophic species, the SCM was elaborated to include intermediates with autoinhibition functions (e.g., peroxides as respiratory by-products) and the possibility of direct incorporation of deficient C-substrate to a pool of reserved compounds (Figure 7). The main feature of growth control in oligotrophic species is the relative autonomy/independence of several metabolic domains: transport, respiration, and biosynthesis.

Many *L*-selected microorganisms (bacilli, actinomycetes, some fungi) share the following common

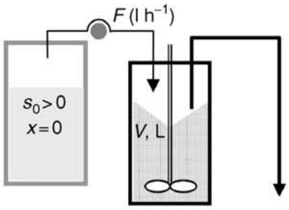
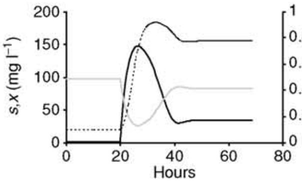
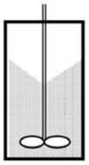
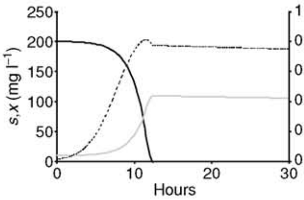
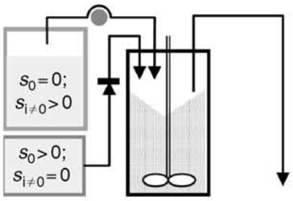
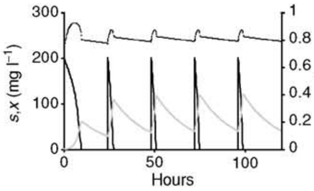
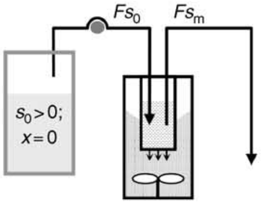
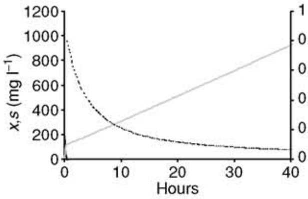
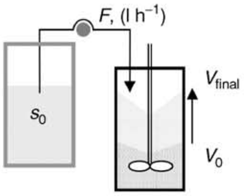
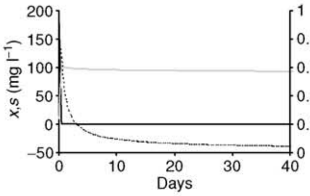
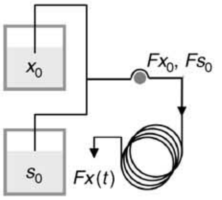
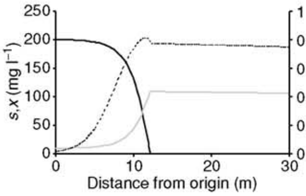
Cultivation method	*Design*	*Equations of mass balance*	*Growth dynamics*
Chemostat		$\frac{\delta s}{\delta t} = D(s_0 - s) - q_s x$ $\frac{\delta x}{\delta t} = \mu x - Dx$ $\frac{\delta r}{\delta t} = \mu\left(\frac{s}{K_r + s} - r\right)$ [12] $q_s = rQ\frac{s}{K_s + s}$; $D = F/V$ $\mu = Y[q_s - mr]$	
Batch		Dilution rate $D = 0$ and $s = s_0$ at $t = 0$; the rest is according to eqn [12]	
Phased culture		D = constant $s_0 > 0$ at $t = i\Delta t$, where $i = 0, 1 \ldots n$, and $s_0 = 0$ at other t; the rest is according to eqn [1]	
Continuous dialysis		$\frac{\delta s}{\delta t} = \frac{F}{V_m}(s_0 - s_m) - q_S x$ $\frac{\delta x}{\delta t} = \mu x$ [13] The rest is according to eqn [12]	
Fed-batch culture (FBC)		$D = F/(V_0 + Ft)$ The rest is according to eqn [12]	
Continuous plug-flow		$\frac{\partial s}{\partial t} + f\frac{\partial s}{\partial z} = D\frac{\partial^2 s}{\partial z^2} - q_s x$ $\frac{\partial x}{\partial t} + f\frac{\partial x}{\partial z} = D\frac{\partial^2 x}{\partial z^2} - \mu(s)x$ The rest is according to eqn [12]	

Figure 5 The major cultivation systems relevant to the purposes of microbial ecology. s, x, and r, limiting substrate concentration, biomass, and variable of physiological state; q_s, specific rate of substrate uptake; mr, maintenance requirement; m, specific growth rate; V, reactor volume; z, distance along tubing reactor in the plug-flow system.

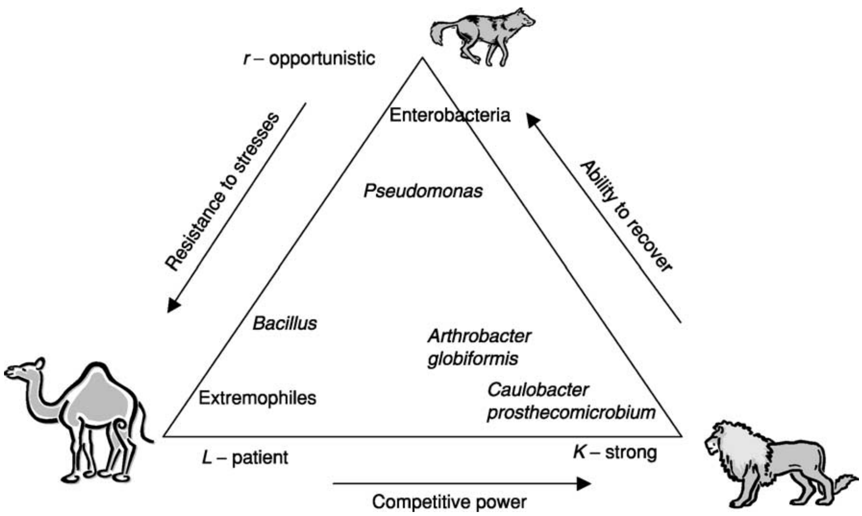


Figure 6 The concept of life strategies. All soil microorganisms are located along three axes characterizing the survival triad: the ability to compete for resources (*K*-axis), recover after stresses (*r*-axis), and resist unfavorable environment (*L*-axis). One can distinguish the following three types of natural selection which correspond to three types of life strategy:

1. *K*-selection operates in climax ecosystems under stable and predictable conditions with no or rare perturbations and stresses. The products of *K*-selection are highly competitive species with 'lion' type of life strategy. The bacterial 'lions' are oligotrophic species that have evolved powerful transporters to absorb the nutrients from diluted solutions (K_s in the range 10^{-6}–10^{-9} mol l^{-1}).
2. *r*-selection operates on the pioneer stages of succession initiated by some perturbation of a climax ecosystem (fire, nutrients flash, cataclysmic elimination of competitors) that has resulted in a temporary relief from competition for nutrient resources. *r*-selected species survive in ephemeral habitats due to mobility and high reproduction rates (opportunistic or 'jackal' type of strategy).
3. *L*-selection operates under adverse environmental conditions caused by abiotic stresses (high/low salt concentration, temperature, pH, water content) or biotic negative interactions (antagonism, starvation caused by depletion of substrate by more successful competitors). The products of *L*-selection are the patient species resistant to a particular stress factor ('camel' type of strategy).

features: endospore formation, production of antibiotics, and synthesis of hydrolytic enzymes. All these features help them to survive even if they lose in direct combat with competitors for deficient nutrients. Kinetic studies allow us to understand why they are weak competitors. The most striking feature was observed in *Bacillus* dialysis culture: the bacteria stopped growing after 2–3 weeks when the residual glucose level dropped below the threshold value of 20–50 μg/l^{-1}. In rich environments, even with an intermittent supply of nutrients (feast-to-famine transitions simulated in batch culture), these bacteria perform well. They rapidly deplete the available substrates, which triggers sporulation and transition to a dormant state, preserving the bacteria from extinction. However, the chronic starvation typical of most oligotrophic environments is the 'trap' for bacilli; they are provoked to sporulate but are not able to finish it in a normal way. The slow feed via the dialysis membrane provides glucose levels too high to allow termination of normal sporulation and too low for growth due to the uncoupling action of metabolite H and acceleration of turnover rate (Figure 7).

Decomposition Kinetics

Decomposition of organic matter is the most essential soil process. There are several ways to quantify kinetics of decomposition.

Decomposition as Self-Decay

The simplest way is to use a first-order kinetic equation:

$$\frac{\delta s}{\delta t} = -ks \qquad [14]$$

where k is a constant dependent on environmental factors. To account for the heterogeneity of detritus, the same equation is written for n (usually $n = 2$ or 3) arbitrarily selected organic matter fractions and then summed as follows:

$$\frac{\delta s}{\delta t} = -(k_1 s_1 + \ldots + k_n s_n) = -\sum_{i=1}^{n} k_i s_i \qquad [14a]$$

Literally, eqn [14a] stands for spontaneous (without involvement of soil organisms or enzymes)

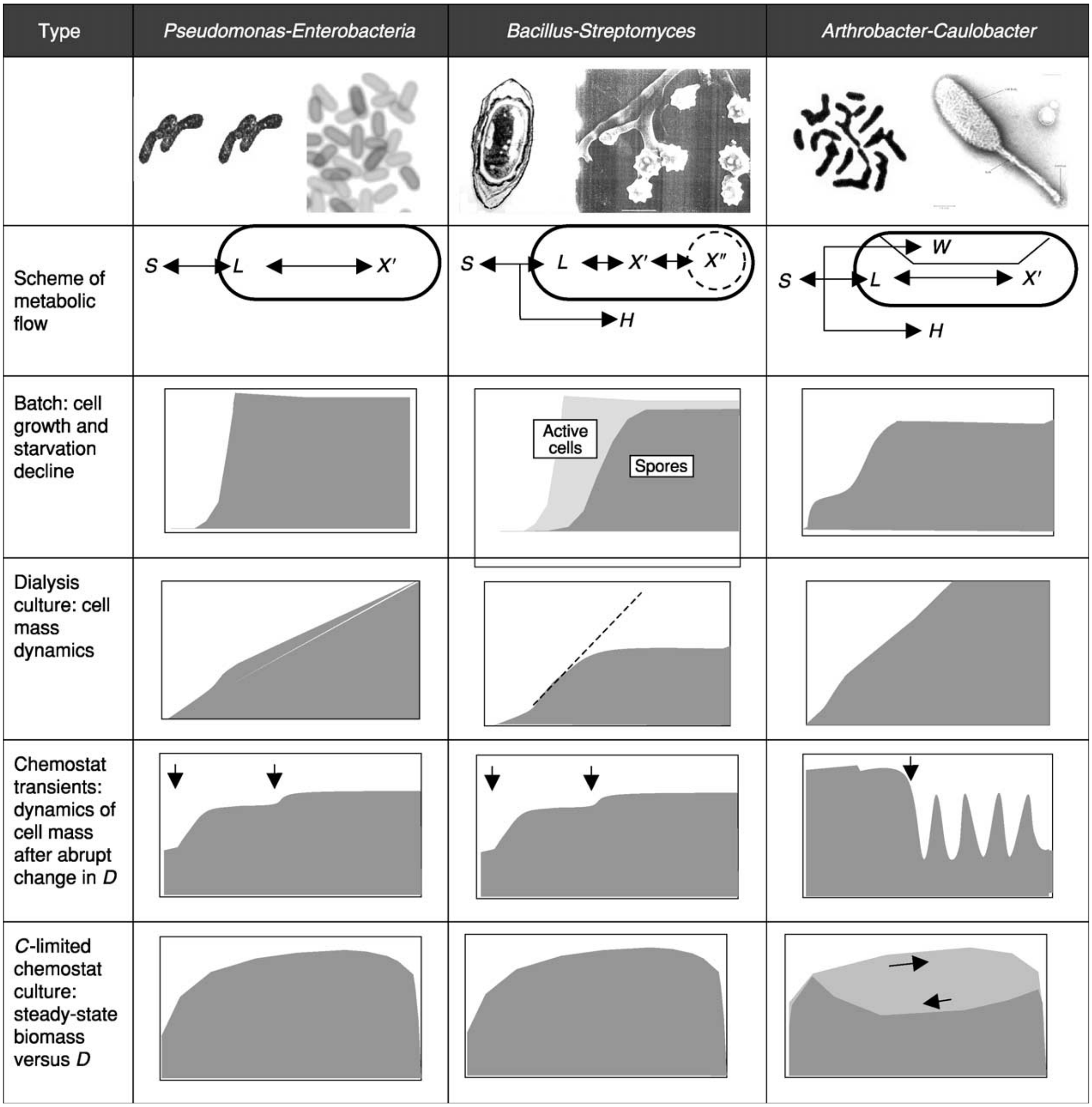


Figure 7 Diversity of growth patterns displayed by soil bacteria in various life strategies in several contrasting types of laboratory cultivation (1) and dialysis (2) cultures as well as chemostat during transient (3) and under steady-state conditions (4). Upper row shows typical morphology of selected microbial groups; next row demonstrates cell growth chart and the main state variables of the respective modification of SCM used for simulation (X'' is prospore compartment, H is autoinhibitor, and W is poly-b-oxybutyrate – reserved compounds): for S, L, and X'. see **Figure 2**. Arrows indicate the time of switching from one dilution rate, D, to another in chemostat culture (3) or derection of sequential D-changes (4, from low D to high or reverse) in chemostat culture displaying multistability.

decomposition of organic matter as well as independent and parallel decay of different individual compounds; no transformation reactions are assumed. All the listed assumptions are unfair; but eqn [14a] is the most popular in modeling of the soil decomposition due to its simplicity.

To account for the transformations accompanying decomposition of organic matter in the soil, e.g., the synthesis of microbial cells and exometabolites from original plant litter, condensation reactions, and formation of organomineral complexes, eqn [14] is modified as follows:

$$\frac{\partial f(z,t)}{\partial t}+\frac{\partial V(z,t)}{\partial z}=-U(z,t)f(z,t) \qquad [15]$$

Here $f(z,t)$ is the distribution of detritus components with respect to their susceptibility to microbial degradation, z, where z ranges between 0 and the maximum value z_m; $V(z,t)$ is the amount of detritus converted into more or less available compounds; and $U(z,t)$ is the degradation rate.

Decomposition of Simple Individual Compounds Added to Soil

More realistic models explicitly account for the soil organisms responsible for biodegradation. Fortunately, such a task does not require a mandatory microbiological analysis, because microbial growth can be evaluated indirectly, from the decomposition dynamic patterns. Such an approach is especially useful in soil incubation studies, when decomposition of added individual organic compounds (pollutants, ameliorators, model compounds) is recorded as residual concentration of chemicals and/or as the CO_2 formation rate. As shown in Figure 8, nonmicrobial degradation should follow zero- or first-order kinetics with a monotonous decline of soil respiration. If enzymes or nongrowing cells are involved, then decomposition follows the Michaelis–Menten equation with its mixed order. The growth of degraders is manifested as a temporary increase in the degradation rate. Overall, more biomass (x_0) before amendment and a faster growth (higher μ) leads to a steeper response and shorter lag-phase. The approximation of substrate-induced soil respiration dynamics with SCM allows us to find the exact value of x_0 and r_0, characterizing the physiological state of studied microorganisms *in situ*.

As another example, decomposition of glucose in bare soil kept for a long time without plants or fresh litter always shows a characteristic wavy dynamic pattern (Figure 8, line 5): rapid glucose uptake the first 1–3 h, followed by deceleration and then acceleration of glucose uptake until all the added compound is consumed. In this case, the respiration dynamics remains monotonic. A remarkably similar pattern was found in a batch culture of *Arthrobacter globiformis* and other *K*-selected oligotrophic species; the first phase of intensive uptake is attributed to filling of storage cell compartment, while the second one is due to normal exponential growth.

Decomposition of Cellulose

Significantly more complicated is the decomposition kinetics of polymeric plant material, e.g., cellulose. In contrast to monomeric substrates, plant polymers are degraded by extracellular enzymes, and the depolymerization products are shared by the whole community. In addition, senescent plant tissues are poor in biogenic elements (N, P, Mg, Fe), while production of extracellular enzymes requires extra amounts of N, Mg, Ca, and P (direct incorporation into protein molecules and support of ribosomal machinery). The deficiency of biogenic elements coupled with C-sufficiency provides a higher risk of metabolic imbalance: catabolic repression of enzyme synthesis, accumulation of inhibitory products, and 'substrate-accelerated death.' In the sphagnum bog, these obstacles are expressed in the highest degree. In addition to metabolic imbalance, these systems are partially anoxic and decomposition rates are so low that peat accumulates. A microbial decomposition under extremely unfavorable conditions is made possible only by the 'cooperative efforts' of the entire community.

The minimal structure of the cellulolytic community consists of three microbial groups: (1) *L*-selected hydrolytic populations producing extracellular cellulases; (2) *r*-selected copiotrophic organisms capable of explosive growth and some way to cope with deficiency of mineral nutrients (e.g., N_2 fixation); and (3) the *K*-selected oligotrophic populations capable of scavenging monomeric products diffusing from decomposition hot spots. In addition, free cellulases immobilized on soil particles provide some background level of depolymerization activity.

The chain of events after cellulose is added to soil is as follows:

1. induction of microbial cellulase synthesis, splitting polymers into mono- and oligosaccharides which are accumulated within the hot spot of microbial activity
2. sugar uptake and ephemeral microbial growth until one of the nutrients, for example N, becomes depleted
3. N-starvation stops hydrolytic activity by driving these organisms into a dormant stage (most hydrolytic microbes form spores), or possibly switching to N_2 fixation
4. growth of this group is stopped when the available C-source is depleted and double N and C starvation occurs, then the third group of *K*-selected oligotrophs continue to function as a sink for depolymerization products (generated by free enzymes).

The oligotrophic bacteria are not affected by N-starvation due to location outside the decomposition hot spot. This verbal description is translated into a mathematical model which gives fair agreement with observations (Figure 9). Fertilization of soil with NPK considerably stimulates cellulose decomposition; however, the highest activity is achieved if soils are air-dried and rewetted before incubation (compare the left and right parts of Figure 9). The

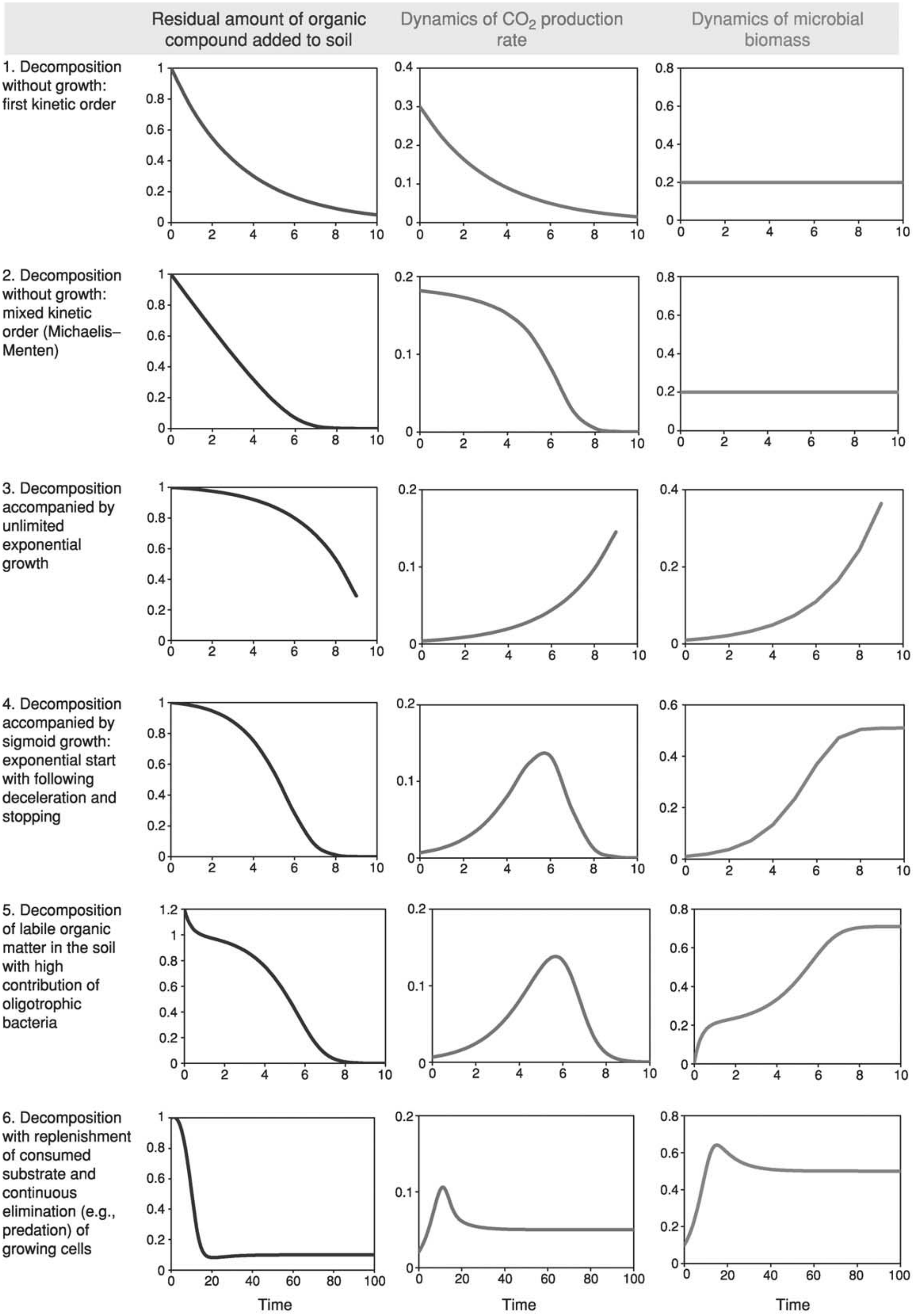


Figure 8 Biodegradation dynamics recorded in soil incubated with added organic substrates. In typical soil incubation studies degradation of labile organics (glucose, amino acids, degradable pesticides, etc.) is followed by recording of residual amount of added compound (first column), CO_2 evolution rate (second column) and microbial biomass (third column). The shape of dynamic curves is a sensitive indicator of the biodegradation mechanisms as indicated by six shown examples.

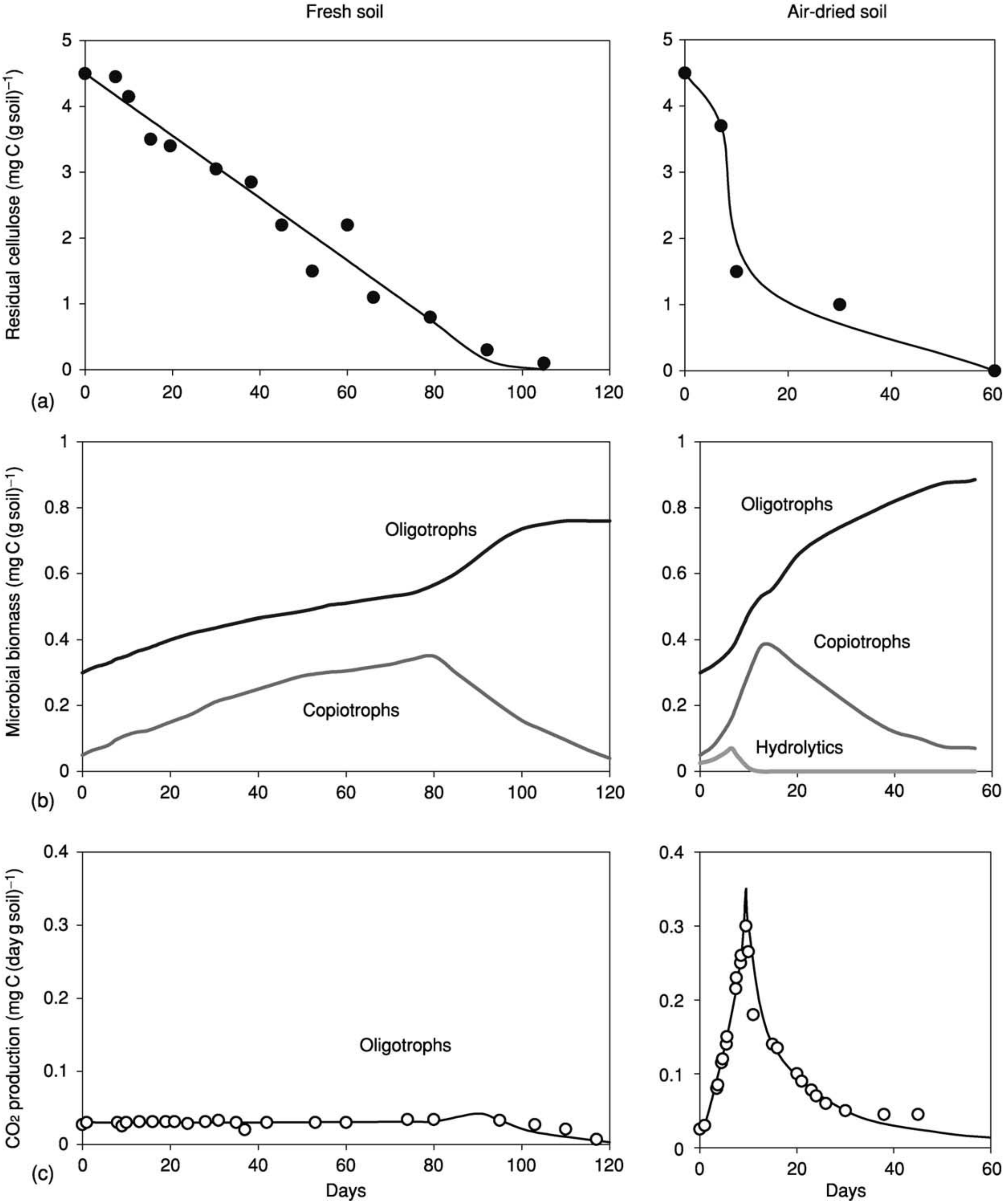


Figure 9 Cellulose decomposition in the soil as related to population dynamics of cellulolytic microbial community. Left: fresh forest soil; right: the same soil after air-drying and rehydration 5 days before cellulose amendment. (a) Residual cellulose; (b) simulated dynamics of: (1) hydrolytics producing extracellular cellulases; (2) copiotrophic N_2-fixing organisms; and (3) the oligotrophic populations capable of scavenging monomeric products diffusing from decomposition hot spots; (c) respiration dynamics. All curves}were simulated by a synthetic chemostat model-based model. Source: Panikov NS (1995) *Microbial Growth Kinetics.* London: Chapman & Hall.

drying–rehydration causes partial damage to microbial cells (similar to fumigation) and release of cell constituents into soil solution. In addition, this treatment removes inhibitors and signaling metabolites, which arrest the biosynthetic activity of hydrolytic microorganisms. It is interesting that healthy soils have restricted hydrolytic activity (slow zero-order decomposition due to action of 'free' soil enzymes). An excessive activation of soil hydrolases takes place more often in pathological cases, e.g., in arable soil after continuous use of NPK and development of soil toxicity.

Decomposition of Plant Litter

The final step from cellulose to lignocellulose and entire plant litter adds the following complications: (1) in lignocellulose the polysaccharide component is further preserved from microbial enzyme by lignin; (2) the heterogeneity of plant litter as a substrate is essentially higher due to the presence of the complex mixture of heteropolysaccharides, waxes, and monomeric phenolic compounds (often considered as inhibitors of decomposition process); and (3) a more complex soil community. However, degradation of cellulose and lignin is carried out jointly, the carbohydrate component being used as available C-source and to generate peroxides driving the oxidative destruction of lignin. Although lignin slows the decay rate, lignocellulose is decomposed as a whole, and there is no significant enrichment of lignin in the old residual substrate. Therefore the kinetic mechanisms of decomposition revealed for cellulose alone are also applicable for lignocellulose.

See also: **Microbial Processes:** Community Analysis; **Mycorrhizal Fungi**

Further Reading

Brock TD (1971) Microbial growth rates in nature. *Bacteriological Review* 35: 39–58.

Gause GF (1934) *The Struggle for Existence*. Baltimore, MD: Williams & Wilkins. Available online at www.ggause.com/Contgau.htm.

Gerson U and Chet I (1981) Are allochtonous and autochthonous soil microorganisms *r*- and *K*-selected? *Revue d'Ecologie et de Biologie du Sol* 18: 285–289.

Hattori T (1973) *Microbial Life in the Soil; An Introduction*. New York: Marcel Dekker.

Lynch JM (1983) *Soil Biotechnology: Microbial Factors in Crop Productivity*. Oxford: Blackwell.

Lynch JM (1990) *The Rhizosphere*. Chichester: Wiley.

Panikov NS (1995) *Microbial Growth Kinetics*. London: Chapman & Hall.

Paul EA and Clark FE (1996) *Soil Microbiology and Biochemistry*. San Diego, CA: Academic Press.

Pirt SJ (1975) *Principles of Microbe and Cell Cultivation*. Oxford: Blackwell Science.

Poindexter JS (1981) Oligotrophy. Fast and famine existence. *Advances in Microbial Ecology* 5: 63–89.

Winogradsky SN (1949) *Microbiologie du Sol*. Paris: Masson.

MICRONUTRIENTS

L M Shuman, University of Georgia, Griffin, GA, USA

Introduction

Micronutrients in soils and their plant availability are critical to plant nutrition. The supply of nutrients to plants is included in the discipline of soil fertility. Micronutrients are elements found in plants at concentrations less than 100 mg kg^{-1} and are generally found at low concentrations in soils, with the notable exception of Fe. Four are present as cations (Zn, Cu, Mn, and Fe), one is an uncharged molecule (B), and one is an anion (Mo) in soil solution and in plants. Several other elements have been shown to be essential to plant nutrition, such as Ni and Cl, but they are not controlled through soil amendments, and deficiencies have only been found in research settings. Not only are the element concentrations important to soil fertility, but more so are the soil properties and processes that influence availability to plants.

Zinc

Zn is taken up by plants in the divalent form, which is the form most often found in soil solution at normal soil pH values. Concentrations in the soil solution that are usually adequate for plant growth are 0.05–0.25 mol l^{-1}, but will become toxic at concentrations greater than 3–6 mol l^{-1}. Soil-solution concentrations of Zn^{2+} are usually low at 0.01–1.0 mol l^{-1} in calcareous soils, but higher in acid soils. There is still uncertainty about whether Zn uptake by plants is an active or passive process. The main interaction with other micronutrient ions is with Cu^{2+}, which can inhibit Zn uptake as well as compete for transport sites within the plant. Zinc interacts with the macronutrient P, in that high P availability can cause Zn deficiency. It is still not known for certain whether the P–Zn interaction takes place mostly in the soil or the plant. Divalent cations such as Mg^{2+} and Ca^{2+} can also interfere with Zn uptake by competition. Zinc deficiencies increase when the soil is waterlogged as for rice-growing. These deficiencies could come from the precipitation of Zn or by increases in Zn–organic complexes.

Zinc is found at higher concentrations in basic rocks than in acidic or sedimentary rocks such as limestone and occurs naturally as sulfides, carbonates, and silicates. The mean concentration in the Earth's crust is 80 mg kg^{-1}. As for most elements, Zn is found at higher concentrations in the topsoil than the subsoil, because plant roots bring Zn to the surface. Of all the soil properties affecting Zn availability to plants, pH is certainly the most important. There are several causes for the increase in availability with decreasing pH: first, the adsorption of Zn by colloids increases with increasing pH; second, Zn can precipitate at high pH values as in calcareous soils. The availability is very low at pH >7.0; finally, diffusion to the plant root is decreased by increasing pH. Zinc usually does not become toxic unless the pH is low (less than 4.0) or the plants are susceptible, as in the case of peanuts. Soil organic matter also has a profound effect on Zn availability. The two forms of organic matter, solid and soluble, have opposing effects on Zn, as they do for most metals. Soluble organic matter–Zn complexes in solution assist in bringing Zn to plant roots, thus enhancing availability. Solid organic matter provides negative adsorption sites, as well as specific adsorption sites that bind Zn and decrease plant availability. As a result, Zn availability is low in organic soils; however, the action of soil microbes in breaking down solid soil organic matter releases Zn and increases plant uptake. Soil secondary clay minerals as well as Fe and Al oxides also have high bonding energies for Zn and high adsorption capacity. These can bind Zn and hinder plant uptake. Zinc is not susceptible to oxidation and reduction in soils, so redox status per se does not affect availability. In soils that undergo oxidation–reduction cycles, the freshly precipitated amorphous (or microcrystalline) Fe oxides are in abundance, and these have approximately 10 times more adsorptive capacity than aged, crystalline Fe oxides.

Copper

Cu is adsorbed by plants as the divalent ion Cu^{2+}. Of the micronutrients considered here, it is most strongly bound with organic chelating materials ($Cu > Zn > Mn > Fe$) and thus is strongly associated both with solid soil organic matter and with soluble organics in soil solution. Up to 98% of Cu found in soil solution is bound by low-weight organic compounds. For adequate growth and development, plants need at least 0.02 mmol l^{-1} in solution. Since the free Cu^{2+} ion is the form taken up by plants, Cu bound by organics is not readily available. Soil moisture has little effect on uptake except that flooding decreases both water-soluble Cu and Cu available to plants.

More Cu is found in soils developed from basic rocks and shales than from sands, sandstones, and acid igneous rocks; basalts are higher in Cu, while granites are low in Cu. Cu deficiencies are prevalent in organic soils such as peats and mucks, and the mean total Cu found in soils is 70 mg kg^{-1}. Cu is strongly associated with oxides of Fe and Al according to adsorption and fractionation studies. As with most metals, it is more strongly adsorbed by soil colloids at higher pH values, but plant availability is not as sensitive to pH as for some other metals because of Cu–organic complexing. Soil-solution Cu concentration in terms of chemical activity is a linear function of soil pH (Figure 1). Plants can develop Cu toxicity symptoms when grown in acid soils of pH less than 6.5. High organic matter content can lead to Cu deficiencies due to the formation of insoluble organic complexes. On the other hand, in calcareous soils, where high pH could decrease Cu availability, soluble organic complexes can help to keep availability above deficient levels. The soluble organics are evidently able to shift the equilibrium from the solid organic complexes to the soluble, so that even at higher pH values the soluble organics can enhance Cu concentrations in solution. Also, over time, the Cu associated with solid organic matter is released through microbial breakdown of the organic matter. Cu does not change oxidation values readily in soils, so the decrease in availability in flooded soils could come from the chemical reduction of crystalline Fe and Mn oxides to amorphous forms, which provide surfaces for Cu adsorption.

Manganese

Mn and Fe are considerably different in their soil chemical reactions from Zn and Cu, because they

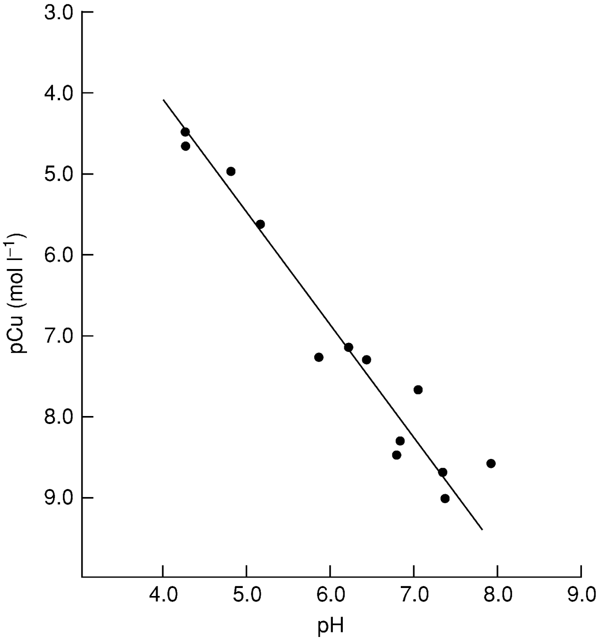


Figure 1 Activity of Cu^{2+} versus soil-suspension pH. pCu, negative $\log_{10}$ of Cu concentration in solution. (Adapted from McBride MB and Blasiak JJ (1979) Zinc and copper solubility as a function of pH in an acid soil. *Soil Science Society of America Journal* 43: 866–870, with permission.)

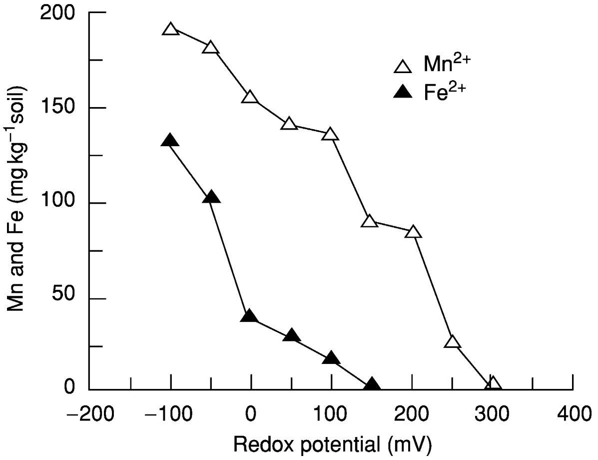


Figure 2 Soluble Mn^{2+} and Fe^{2+} in soil suspensions versus redox potential during an oxidized-to-reduced cycle. (Adapted from Patrick WH Jr and Jugsujinda A (1992) Sequential reduction and oxidation of inorganic nitrogen, manganese, and iron in flooded soil. *Soil Science Society of America Journal* 56: 1071–1073, with permission.)

are easily oxidized and reduced in soils. Mn is taken up by plants as Mn^{2+}, although most of the Mn in soils is in the Mn^{4+} oxidation state. Mn concentrations in soil solution can vary from less than 0.1 mmol l^{-1} in well-aerated soils to as high as 400 mmol l^{-1} in submerged soils. Plants need a range of 0.01 to approximately 50 mmol l^{-1}. Chelation of Mn occurs in soils as for the other metals, but the uptake of chelated Mn by plants is poorly understood. High soil-moisture content tends to reduce the oxidized Mn^{4+} form to the reduced Mn^{2+} form, making it more available. Thus, deficiency is not a problem on wet soils, and toxicity may result, especially for susceptible plants.

Mn is the 11th most common element in the Earth's crust and averages 0.09% or 900 mg kg^{-1}. The oxides or sulfides are often associated with Fe. Mn can be found in soil minerals such as MnO_2, MnOOH, and Mn_3O_4. Mn is low in soils that are derived from crystalline shales and acid igneous rocks, but is higher in soils derived from basalt, limestone, and shales, especially limestones. Deficiencies occur in organic, calcareous, and poorly drained acid sandy soils. Usually the higher the clay content of a soil, the greater will be the Mn content. For example, the soils of the US southeastern piedmont have significant amounts of clay and some are derived from limestone. These soils are so high in Mn that toxicities can result. However, on the southeastern lower coastal plain, the sandy soils are alternately wet and dry and can cause Mn deficiencies. Manganese availability is tied closely to the oxidation–reduction status of soils. Soils that are waterlogged are reduced and release Mn^{2+} from the oxidized forms. The more negative the redox potential, the higher the Mn concentration in soil solution (Figure 2).

Besides oxidation–reduction, the other major influence on Mn availability is soil pH. At pH values greater than 6 and under high-oxidation conditions, Mn is mostly in the insoluble Mn^{4+} form. At lower pH and/or under wet conditions, the Mn is in the soluble Mn^{2+} form and is readily available to plants. At pH values less than 5, oxidation–reduction is not as important as pH in determining Mn availability, but at pH values of 6–8, oxidation–reduction becomes very important. Plant roots themselves exude protons and chelating agents which can assist in mobilizing Mn. Solid soil organic matter can complex Mn with stability increasing at increasing pH.

Iron

Approximately 5.1% of soil on the average is Fe, but availability can be low in well-aerated soils. Soil Fe is derived from ferromagnesium silicates, which includes a great number of rock types. Mineral forms are oxides, hydroxides, and silicates, as well as amorphous oxides. These minerals take on forms such as goethite and hematite depending on the hydration status and crystalline form. The Fe released by weathering is redistributed into secondary minerals. The solubility of soil Fe depends very much on pH. Iron is solubilized from solid Fe oxides and hydroxides by the following reaction with protons:

$$\text{Soil Fe} + 3H^+ = Fe^{3+} + 3H_2O \qquad [1]$$

In solution, Fe hydrolyzes to Fe^{3+}, $Fe(OH)_2^+$, and $Fe(OH)_4^-$, which is again dependent on the pH. As for Mn the oxidation–reduction status of the soil also affects Fe solubility. Figure 2 shows that, as for Mn, the more positive the redox potential, the less the Fe concentration in soil solution. The divalent form is more soluble than the trivalent form and is the form that plants absorb. But in well-aerated soils at pH values of approximately 6–8, the Fe is in the Fe^{3+} state and is relatively insoluble, with concentrations around 10^{-8}–10^{-3} mol l^{-1}. Submergence of soils can cause chemical reduction of Fe^{3+} to Fe^{2+}, resulting in Fe toxicity. However, damp soils can cause a decrease in the Fe absorption capacity of plant roots and affect bicarbonate in calcareous soils, decreasing uptake. Fe is strongly complexed by soil organic matter, and the soluble organic matter can increase the soil-solution Fe concentrations to 10^{-5}–10^{-3}. This soluble organic matter comes from solid organic matter degradation and from release of siderophores (nonprotein complexes) by plant roots and soil bacteria and fungi. Plants need Fe concentrations of the order of 10^{-4}–10^{-2}, so plants must have mechanisms to increase soil Fe availability. Plants have two strategies to increase Fe availability. What has been named strategy I is release of reductants to reduce Fe^{3+} to Fe^{2+}, and strategy II is release of Fe-binding organic ligands and protons to help solubilize soil Fe. Strategy II is used by grasses and strategy I by dicotyledons and nongraminaceous monocotyledons.

Boron

B and Mo are found in combination with oxygen in nature. A common mineral for B is borax, with the formula $Na_2B_4O_7 \cdot 10H_2O$. B is present in soil solution and in plants as boric acid ($B(OH)_3$), a compound that is considered an acid because it hydrolyzes according to the following reaction, producing hydronium ions:

$$B(OH)_3 + 2H_2O = B(OH)_4^- + H_3O^+ \qquad [2]$$

This ionic form is a very weak acid, which accepts OH^- ions; at usual soil and plant pH values, the acid is undissociated. The mechanism of uptake is not well understood. B has a narrow range of sufficiency to toxicity for plants. Plants need 0.01 mmol l^{-1} in soil solution, but 0.03 mmol l^{-1} is toxic to B-sensitive plants, which include citruses and peaches. Drought conditions in soil increase B deficiency in many crops, which was first noticed in sugar beet; B deficiency restricts root growth, having more severe effects than any of the other micronutrients. Drought also restricts the mineralization and availability of organically bound B.

One mineral that is known for its B content is tourmaline; however, it is so insoluble that it cannot be considered a B source for plants. Boron is found as a substitute for Si in other minerals, and the rate of weathering of these minerals is slow. Soils derived from granite and other igneous rocks are usually low in B content, whereas those derived from sedimentary rocks of marine origin are usually higher in B content. The mean concentration of B in the crust is 50 mg kg^{-1}. B is water-soluble and is easily leached from soils, so sandy, porous soils can often be B-deficient. Sandy soils have more B available to the plant at a given hot water-extractable level than finer-textured soils, because the B is bound more strongly by the clays and oxides in the finer-textured soils (Figure 3). B is sometimes toxic in arid and semiarid soils; the B content in irrigation water can cause B buildup in semiarid soils, leading to toxicity. As for most other micronutrients, high soil pH (6–9) causes low B availability. Higher pH levels mean higher adsorption on soil clays and oxides, thus liming can have a large effect on B availability. At lower pH values, the uncharged boric acid is not readily adsorbed and is held by ligand exchange; at higher pH values, the $B(OH)_4^-$ ion forms and is strongly adsorbed by electrostatic anion exchange. At low pH, soil B can become toxic to plants. Most B in the soil is contained in the organic matter and becomes more available as the conditions for microbial degradation become favorable, such as warmer temperatures, higher moisture, and better aeration. In organic soils such as peats and mucks, the availability can become too high as the B in the organic matter is

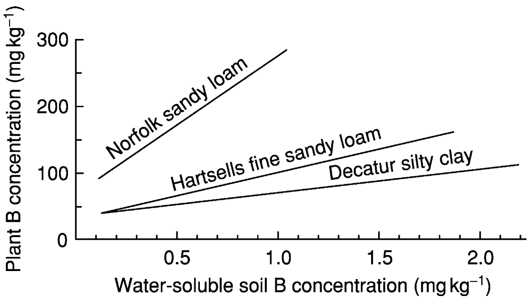


Figure 3 Plant boron (B) concentration versus water-soluble soil B concentration for three soils of different textures. (Adapted from Wear JJ and Patterson RM (1962) Effect of soil pH and texture on the availability of water-soluble boron in the soil. *Soil Science Society of America Proceedings* 26: 344–346, with permission.)

mineralized from the microbial biomass. B does not undergo oxidation–reduction reactions in soils.

Molybdenum

Mo concentrations are the lowest in soils of any of the micronutrients, being only 2.3 mg kg^{-1} on average in the crust. Soil Mo is derived more from shale and granite rocks, which are higher in content than sandstones or limestones. Soils that are very high in organic matter content, are alkaline, or are poorly drained are generally higher in Mo availability than other soils. Mo is different in its response to soil pH from the other micronutrients. As soil pH increases, the Mo plant availability also increases (Figure 4). Mo is adsorbed by Fe and Al oxides as the molybdate ion, MoO_4^{2-} at pH > 4.2, and this adsorption decreases at higher pH values. Thus, liming soils increases the uptake of Mo. Mo toxicity is rare in agricultural crops, but has been observed in controlled experiments and on peats and mucks that are poorly drained. Mo availability is higher in soils as the organic matter content or the Fe and Al oxide content decreases. Since Mo is so strongly associated with Fe and Al oxides through adsorption and precipitation, Mo availability is negatively correlated with the contents of Fe and Al oxides, whereas total Mo content may be positively correlated with Fe and Al oxide content. Oxidation and reduction processes do not affect Mo, so any oxidation–reduction effects on availability are indirect. Mo is needed by plants for nitrate reduction and symbiotic N-fixation by legumes. Deficiencies are usually on acid soils and are usually in legumes. Increases in soil moisture increase Mo availability due to decreased soil adsorption of MoO_4^{2-} under the higher soil pH conditions resulting from waterlogging. The reducing conditions caused by waterlogging reduce Fe^{3+} to Fe^{2+}, and Mo is less associated with the reduced form of Fe so therefore more available. Also, under wet conditions, Mo moves to the plant root more by mass flow than it does under drier conditions.

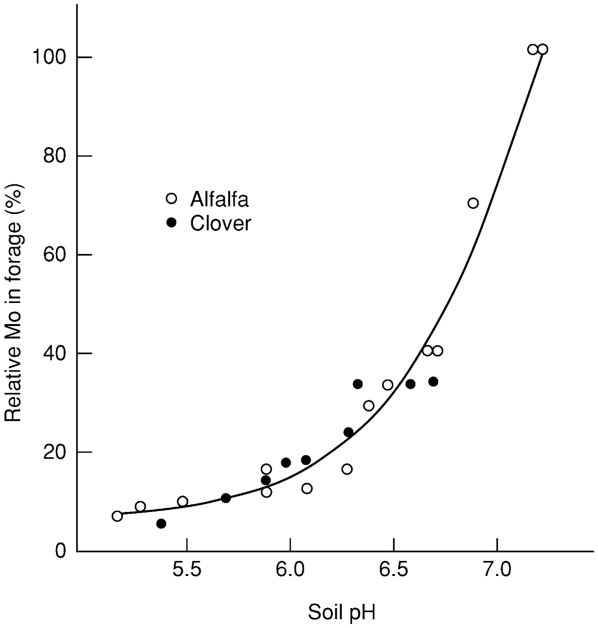


Figure 4 Molybdenum in forage as a percentage of maximum versus soil pH. (Adapted from Mortvedt JJ and Anderson OE (1982) Forage legumes: diagnosis and correction of molybdenum and manganese problems. Southern Cooperative Series Bulletin 278. Athens, GA: University of Georgia, with permission.)

Micronutrient Soil Tests

This discussion will be limited to soil chemical testing, which is the major diagnostic tool used for both macronutrients and micronutrients, designed to predict micronutrient availability to the crop. If the soil is found to be nutrient-deficient, the test should give an indication of the severity of the deficiency and possible yield losses. Field calibration studies are carried out to estimate the amounts of micronutrient fertilizers to be added in response to a certain soil-test result. In interpreting soil-test data, a 'critical level' is used below which are expected yield responses to micronutrient fertilizers. One approach is the Cate–Nelson procedure, where a scatter diagram is formed from soil-test results on the x-axis and yields on the y-axis. Crossed lines are drawn such that a minimum number of points are in the upper left and lower right quadrants (Figure 5). The value found on the x-axis

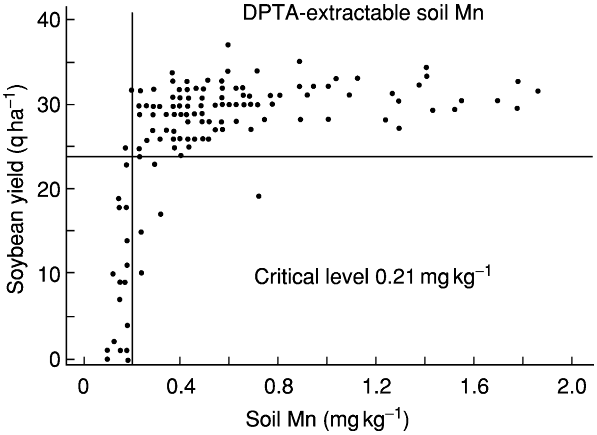


Figure 5 Cate–Nelson plot for soybean yield versus diethylenetriaminepentaacetic acid (DTPA)-extractable soil Mn. (Adapted from Shuman LM, Boswell FC, Ohki K, Parker MB, and Wilson DO (1982) Critical soil manganese deficiency levels for four extractants for soybeans grown in a sandy soil. *Soil Science Society of America Journal* 44: 1021–1025, with permission.)

is the critical level below which fertilizer is applied to keep yields at maximum. Another approach is correlation analysis. Since some micronutrients are very sensitive to pH, an approach often used is to have different critical levels for various pH levels. At times other soil properties are combined with soil-test results, but pH is usually the only value used.

Micronutrient soil tests should be able to extract, from the soil samples of the micronutrients in question in proportion to those taken up by plants; in other words, the samples should contain micronutrients in the same forms as those available to the plants. Soil is usually sampled in the root zone at approximately 0 to 15–20 cm. Soil tests should be accurate, repeatable, rapid, and correlated well with crop response from severe deficiency to sufficiency. Ideally a test will show differences in availability due to changes in soil properties such as pH or oxidation–reduction conditions. Soil tests are usually designed to remove the soluble and exchangeable forms, but in some instances they dissolve mineral forms that contain precipitated or occluded forms, as for Mo. These chemical extractants are usually fairly simple reagents such as water, dilute acids or bases, and salt solutions, and more recently include chelating agents such as diethylenetriaminepentaacetic acid (DTPA) or ethylenediaminetetraacetic acid (EDTA). Dilute acids commonly in use are 0.1 mol l^{-1} HCl, and an acid mixture of 0.025 mol l^{-1} H_2SO_4 and 0.05 mol l^{-1} HCl (Mehlich 1). These acids are used on acid soils to dissociate, displace, or dissolve the cations from soil clays, oxides, or organic matter. The chelating agent DTPA is found in Mehlich 3, DTPA, and AB-DTPA; and EDTA is in the acid ammonium acetate–EDTA extractant. All these tests extract multielements. Chelates promote extraction from surfaces of clays, oxides, and organic matter. See Table 1 for commonly used micronutrient soil tests, with the deficiency critical ranges for each.

Zn is usually extracted using a multielement extractant, e.g., Mehlich 1 and 3, AB-DTPA, and 0.1 mol l^{-1} HCl. Some of these are used in conjunction with soil pH tests. Cu deficiencies are rare, but Cu is routinely tested using the same extractants as for Zn. Since Cu binds readily with organics, the better extractants are those with chelating agents. For Mn, dilute acids and chelating agents are both used, such as in Mehlich 1 and 3, 0.1 mol l^{-1} HCl, 0.03 mol l^{-1} H_3PO_4, and DTPA tests. Soil pH is often used with these extractants, using a 'sliding scale' for critical levels based on soil pH. Some chelating reagents such as DTPA and AB-DTPA were designed for use with Fe. The dilute acids mentioned are also used for Fe extraction as a soil test. Hot water was found to be a good soil test extractant for B in 1939 and it is still in use today. Mo is associated with oxides and is extracted by acid ammonium oxalate, which also extracts Fe and Al oxides. Soil pH is a big factor in Mo deficiency, so is often used in conjunction with the soil test.

Table 1 Major soil diagnostic tests and critical ranges for deficiency

Micronutrient	*Soil-test method*	*Critical range*
Zinc	AB-DTPA	0.5–1.0 mg kg^{-1}
	DTPA	0.2–2.0 mg kg^{-1}
	Mehlich 1	0.5–3.0 mg kg^{-1}
	Mehlich 3	1.0–2.0 mg kg^{-1}
	0.1 mol l^{-1} HCl	1.0–5.0 mg kg^{-1}
Copper	AB-DTPA	0.5–2.5 mg kg^{-1}
	DTPA	0.1–2.5 mg kg^{-1}
	Mehlich 1	0.1–10 mg kg^{-1}
	Mehlich 3	0.3–15 mg kg^{-1}
	0.1 mol l^{-1} HCl	0.1–2.0 mg kg^{-1}
Manganese	AB-DTPA	0.5–5.0 mg kg^{-1}
	DTPA	1.0–5.0 mg kg^{-1}
	Mehlich 1	5.0 mg kg^{-1} at pH 6 10 mg kg^{-1} at pH 7
	Mehlich 3	4.0 mg kg^{-1} at pH 6 8.0 mg kg^{-1} at pH 7
	0.1 mol l^{-1} HCl	1.0–4.0 mg kg^{-1}
Iron	AB-DTPA	4.0–5.0 mg kg^{-1}
	DTPA	2.5–5.0 mg kg^{-1}
Boron	Hot water	0.1–2.0 mg kg^{-1}
	Mehlich 3	0.7–3.0 mg kg^{-1}
Molybdenum	Ammonium oxalate–pH 3.3	0.1–0.3 mg kg^{-1}

Reproduced with permission from Sims JT (2000) Soil fertility evaluation. In: Sumner ME (ed.) *Handbook of Soil Science*, pp. 113–153. Boca Raton, FL: CRC Press.

Micronutrient Fertilizers

Micronutrient fertilizers include four major types: inorganic salts, synthetic chelates, natural organic complexes, and frits. Certainly the most commonly used are the inorganic sources, since they are the least expensive (Table 2). These include oxides, carbonates, and metallic salts such as the sulfates, chlorides, and nitrates. Manganese oxide and zinc oxide are used to some extent, but are relatively insoluble and become available over a long period of time. The sulfates of Zn, Cu, Mn, and Fe are the most common fertilizers in use, being quite soluble and immediately available to the crop. For B, borax is the fertilizer of choice, and for Mo it is sodium molybdate.

The synthetic chelates include *N*-hydroxylethylenediaminetriacetic acid (HEDTA) and EDTA, with EDTA salts being the most commonly used (Table 3). The rates of chelate application are about two to five times less than inorganic sources, but the cost is 5–100 times greater. The chelates are usually

Table 2 Inorganic fertilizer sources for micronutrient fertilizers

Micronutrient source	*Solubility in water*	*Element concentration (%)*
Zinc		
$ZnSO_4 \cdot H_2O$	Soluble	36
$ZnSO_4 \cdot 7H_2O$	Soluble	22
$ZnCl_2$	Soluble	47
$ZnSO_4 \cdot 4Zn(OH)_2$	Slightly soluble	55
$ZnCO_3$	Insoluble	52
ZnO	Insoluble	60–78
Zinc oxysulfate	Variable	25–60
Copper		
$CuSO_4 \cdot H_2O$	Soluble	35
$CuSO_4 \cdot 5H_2O$	Soluble	20
$CuSO_4 \cdot 3Cu(OH)_2 \cdot H_2O$	Insoluble	37
CuO	Insoluble	75
Manganese		
$MnSO_4 \cdot XH_2O$	Soluble	24–30
$MnCl_2 \cdot 4H_2O$	Soluble	28
$MnCO_3$	Insoluble	31
MnO	Insoluble	41–68
Manganese oxysulfate	Variable	30–50
Iron		
$FeSO_4 \cdot H_2O$	Soluble	33
$FeSO_4 \cdot 7H_2O$	Soluble	20
$Fe(SO_4)_3 \cdot 9H_2O$	Soluble	20
$FeSO_4 \cdot (NH_4)_2SO_4$	Soluble	22
Boron		
$Na_2B_4O_7$ (anhydrous borax)	Soluble	20
$Na_2B_4O_7 \cdot 5H_2O$ (fertilizer borate)	Soluble	14
$Na_2B_4O_7 \cdot 10H_2O$ (borax)	Soluble	11
H_3BO_3	Soluble	17
$Ca_2B_6O_{11} \cdot 5H_2O$ (colemanite)	Slightly soluble	10
Molybdenum		
Na_2MoO_4 (anhydrous)	Soluble	47
$Na_2MoO_4 \cdot 2H_2O$	Soluble	39
$(NH_4)_2MoO_4$	Soluble	49
MoO_3	Slightly soluble	66
$CaMoO_4$	Insoluble	48

Reproduced with permission from Mortvedt JJ, Cox FR, Shuman LM, and Welch RM (eds) (1991) *Micronutrients in Agriculture*. Madison, WI: Soil Science Society of America.

Table 3 Organic fertilizer sources for micronutrient fertilizers

Micronutrient source	*Formula*	*Element concentration (%)*
Zinc		
Zinc chelate	$Na_2ZnEDTA$	14
Zinc chelate	NaZnHEDTA	9
Zinc chelate	NaZnNTA	9
Zinc lignosulfonate	—	5–8
Zinc polyflavonoid	—	5–10
Copper		
Copper chelate	$Na_2CuEDTA$	13
Copper chelate	NaCuHEDTA	9
Copper lignosulfonate	—	5–8
Copper polyflavonoid	—	5–7
Manganese		
Manganese chelate	$Na_2MnEDTA$	5–12
Manganese lignosulfonate	—	5
Manganese polyflavonoid	—	5–7
Iron		
Iron chelate	$Na_2FeEDTA$	5–14
Iron chelate	NaFeHEDTA	5–9
Iron chelate	NaFeEDDHA	6
Iron chelate	NaFeDTPA	10
Iron lignosulfonate	—	5–8
Iron polyflavonoid	—	9–10
Iron methoxyphenylpropane	FeMPP	5

Reproduced with permission from Mortvedt JJ, Cox FR, Shuman LM, and Welch RM (eds) (1991) *Micronutrients in Agriculture*. Madison, WI: Soil Science Society of America.

sold as liquid preparations so are sprayed either on the soil or directly on the crop. Elements applied as chelates are Zn and Fe. Manganese chelate can be sprayed on a growing crop, but, when applied to the soil, the Mn is quickly replaced by Fe and the Mn is ineffective. The high concentrations of Fe in the soil solution cause Fe to replace the Mn on the organic chelate molecule.

Natural organic sources most often in use are the lignosulfonate complexes made from a by-product of the wood-pulping process in paper mills. These complexes are cheaper than synthetic chelates, but are also not as effective because they are readily decomposed in the soil. Inorganic salts are reacted with these organic by-products to form complexes. The elements applied with natural chelates are Zn and Fe.

The last major fertilizer type is frits. These are glass beads with micronutrients occluded in them by melting sand with the inorganic salts added. Like the oxides, these slowly release the micronutrients and are used for maintenance. They are most often applied in sandy soils with a high water table, where leaching of soluble salts is likely to move elements below the root zone before they can be taken up by the plant.

Finally, another source is industrial waste products that contain micronutrients, such as wastes from the steel industry. Even farm manures can be micronutrient sources; for example, pig manure is high in Cu and poultry manure high in Zn.

List of Technical Nomenclature

AB-DTPA	A soil-testing method using a solution of 1.0 mol l^{-1} ammonium bicarbonate (NH_4HCO_3) and the DTPA soil test solution buffered at a pH of 7.6
DTPA	Diethylenetriaminepentaacetic acid

DTPA	A soil-testing method using a solution of $0.005\,\mathrm{mol\,l^{-1}}$ DTPA and $0.01\,\mathrm{mol\,l^{-1}}$ $CaCl_2$ buffered at a pH of 7.3 by $0.1\,\mathrm{mol\,l^{-1}}$ triethanolamine (TEA)
EDDHA	Ethylenediaminedi(*o*-hydroxyphenylacetic) acid
EDTA	Ethylenediaminetetraacetic acid
HEDTA	*N*-Hydroxyethyl-ethylenediaminetriacetic acid
Mehlich 1	A soil-testing method using a solution of $0.0125\,\mathrm{mol\,l^{-1}}\,H_2SO_4 + 0.05\,\mathrm{mol\,l^{-1}}$ HCl
Mehlich 3	A soil-testing method using a solution of $0.2\,\mathrm{mol\,l^{-1}}\,CH_3COOH + 0.25\,\mathrm{mol\,l^{-1}}$ $NH_4F + 0.013\,\mathrm{mol\,l^{-1}}\,HNO_3 +$ $0.001\,\mathrm{mol\,l^{-1}}$ EDTA $+ 0.25\,\mathrm{mol\,l^{-1}}$ NH_4NO_3
NTA	Nitrilotriacetic acid
pCu	Negative $\log_{10}$ of the Cu concentration in solution in moles per liter
pH	Negative $\log_{10}$ of the H^+ concentration in solution in moles per liter

See also: **Fertilizers and Fertilization**; **Iron Nutrition**; **Nutrient Availability**

Further Reading

Adriano DC (1986) *Trace Elements in the Terrestrial Environment*. New York: Springer-Verlag.

Barber SA (1984) *Soil Nutrient Bioavailability*. New York: John Wiley.

Mengel D and Rehm G (2000) Fundamentals of fertilizer application. In: Sumner ME (ed.) *Handbook of Soil Science*, pp. 155–175. Boca Raton, FL: CRC Press.

Mortvedt JJ (2000) Bioavailability of micronutrients. In: Sumner ME (ed.) *Handbook of Soil Science*, pp. 71–88. Boca Raton, FL: CRC Press.

Mortvedt JJ, Cox FR, Shuman LM, and Welch RM (eds) (1991) *Micronutrients in Agriculture*. Madison, WI: Soil Science Society of America.

Rendig VV and Taylor HH (1989) *Principles of Soil–Plant Interrelationships*. New York: McGraw-Hill.

Shuman LM (1998) Micronutrient fertilizers. In: Rengel Z (ed.) *Nutrient Use in Crop Production*, pp. 165–195. New York: Food Products Press.

Sims JT (2000) Soil fertility evaluation. In: Sumner ME (ed.) *Handbook of Soil Science*, pp. 113–153. Boca Raton, FL: CRC Press.

MINERAL–ORGANIC–MICROBIAL INTERACTIONS

P M Huang, University of Saskatchewan, Saskatoon, SK, Canada
M C Wang, National Chung Hsing University, Taichung, Taiwan
M K Wang, National Taiwan University, Taipei, Taiwan

Introduction

Minerals, organic components, and microorganisms are three major solid components of the soil. They profoundly affect the physical, chemical, and biological properties, and processes of terrestrial ecosystems. They should not be considered as separate entities but rather as a united system constantly in close association and interactions with each other in soil environments.

Interactions of these three solid components mediated by soil solution and atmosphere govern mineral weathering transformation, the formation of organo–mineral complexes, microbial and enzymatic activities, soil structural stability and dynamics of aggregate turnover, biogeochemical cycling of C, N, P, and S, and transformation and dynamics of organic pollutants and metals. Therefore, interactions of these components have enormous impacts on terrestrial processes critical to environmental quality and ecosystem health.

Mechanisms of Binding of Organic Matter by Soil Mineral Colloids

Nature of Mineral Colloid Surfaces

The surface properties of soils are dependent on the weathering processes of the parent materials that are initiated by geological, hydrological, and biological agents. Therefore, soil formation processes have a tremendous impact on the formation and modification of surface properties of soils. The surface of 2:1 expansible layer silicates is negatively charged. The ditrigonal cavities of the surface can thus adsorb positively charged ions and polymers to form hydroxy-interlayered minerals in soil and related environments

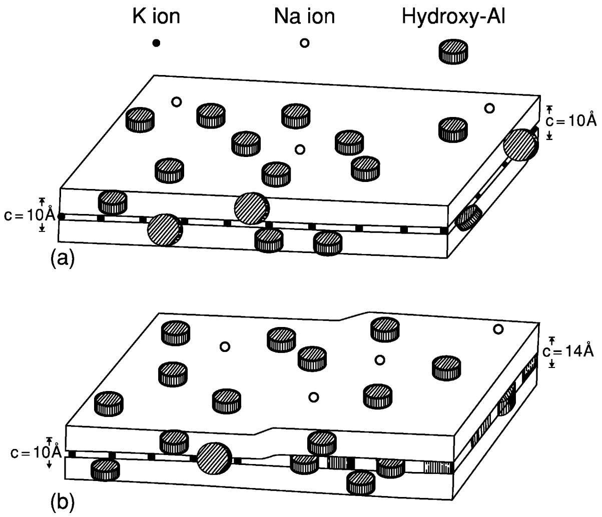


Figure 1 Structural models depicting the hydroxy-Al coatings on external surfaces (a) edge and planar surfaces of micas; and (b) of mica-vermiculite and on interlamellar surfaces of K-depleted layers of mica-vermiculite. Reproduced with permission from Huang PM (1980) Adsorption processes in soil. In: Hutzinger O (ed.) *The Handbook of Environmental Chemistry*, vol. 2/part A. *Reactions and Processes*, pp. 47–59. Berlin: Springer-Verlag.

(Figure 1). These hydroxy-interlayered minerals occur in soils throughout the world and are present in soils of several orders of soil taxonomy systems. In addition, positively charged hydroxy polymers can be adsorbed on the edges and external surfaces of phyllosilicates (Figure 1).

The complexity of surface properties of soil mineral colloids is further manifested by the formation of iron oxyhydroxide or calcium carbonate coatings on the external surfaces of layer silicates, the growth of two-dimensional solid solutions of trace metal oxides on the periphery of Mn oxide nodules, and the coating of calcium carbonate on Mn oxides. Similar effects can occur on the surfaces of other minerals in soils. The surfaces of soil mineral colloids can also complex with organic matter through the formation of inner-sphere and/or outer-sphere complexes. The surface complexation reactions may result in the formation of thick envelopes of colloidal organics on metal oxides and aluminosilicates. Therefore, the surface chemistry of soil minerals is mainly that of the functional groups exposed on the coatings rather than that of the groups in the underlying matrix. The surface chemistry of soil mineral colloids has a very significant effect on the transformation and dynamics of organic matter in soils.

Binding of Organic Matter

Organic molecules interact with the surfaces of mineral colloids by a variety of mechanisms, as summarized below. The reaction mechanisms mainly depend on the intrinsic nature and properties of the organic species, the kind of exchangeable cation on the surface of the mineral, the water content of the system, and the properties of mineral colloids.

Ion exchange Organic cations such as protonated amines and quarternary ammonium cations will enter into exchange reactions with metal cations on the surfaces of mineral colloids. Organic cations are often preferred over metal cations because of size considerations.

Protonation Some organic molecules may become cationic after adsorption at the surfaces of mineral colloids by protonation. The relative basicity of the organic species and the Brønsted acidity of the mineral surface determine whether or not protonation happens. The degree of protonation of the organic molecule is determined by its intrinsic basicity and the ability of the surfaces of the mineral colloids to donate protons. The nature of exchangeable cation, the water content, and the nature of mineral colloids play an important role in influencing the degree of protonation of organic molecules.

Coordination and ion dipole Coordination or ion–dipole interaction is another kind of mineral colloid–organic interaction. Where exchangeable cations are of the transition metal type, unfilled d orbitals will allow the formation of coordinate covalent bonds wherein an electron donor such as N or O on functional groups of organic molecules is available as a ligand. Where exchangeable cations are not able to form classical coordination complexes, they will interact with polar molecules through ion–dipole interaction processes.

Hydrogen bonding Hydrogen bonding is another type of interaction which is possible between organic molecules and mineral surfaces. Where a polar organic molecule cannot displace a water molecule solvating an exchangeable metal cation, it settles for a secondary role by H bonding with the directly coordinated water. This type of arrangement is referred to as 'water bridge,' where water acts as the linking entity between an exchangeable metal cation and a polar organic molecule.

Hydrophobic bonding on clay–organic complexes Natural clay–organic complexes may have adsorption properties quite different from pure clays. Many organic molecules, including aromatics and particularly the halogenated types such as chlorinated and brominated phenyls and biphenyls, are adsorbed a little, if at all, on clay surfaces. The surfaces of clays can be made hydrophobic after reaction with organic molecules which have some hydrophobic properties. When this happens, the surfaces of clay–organic complexes become hydrophobic and in turn organophilic. Therefore, hydrophobic portions of the organic adsorbate can react with the hydrophobic organic portions of clay–organic complexes.

Anion adsorption Although pure clay minerals are negatively charged, hydroxy-interlayered clay minerals have positively charged sites (Figure 1). Organic anions can interact with polyvalent cations such as Al and Fe on the surface coatings of clay minerals.

Adsorption on external mineral surfaces The extent of adsorption of organic matter on external mineral surfaces depends on the physical and chemical characteristics of the surface, the pH of the system, and its water contents. Complexation reactions of humic acid (HA) and fulvic acid (FA) with structural cations of edges of mineral colloids are important binding mechanisms. Hydrogen bonding, which apparently involves H and O of COOH and OH groups of HA and FA and O and H of external planar surfaces and edges, is also of considerable importance.

Adsorption in clay interlayers Interlayer adsorption of FA by expansible silicates is pH-dependent, being greatest at low pH, and no longer occurring at pH 5.0 (Figure 2). An inflection occurs in the adsorption–pH curve near the pH corresponding to the pK of FA. Since the degree of dissociation of the FA is very low at low pH, undissociated FA can penetrate the clay interlayer and displace water coordinated to the cation in the interlayer. Therefore, interlayer adsorption of FA is greatest at low pH. One of the well-defined precursors for the formation of humic substances, hydroquinone, can be transformed in aqueous solution even at near-neutral pH (6.5) to humic macromolecules and adsorbed in the interlayers of nontronite (Fe(III)-bearing smectite) saturated with Ca (Figure 3); this is the most common and most abundant exchangeable cation in soils and sediments. The decrease in the organic carbon contents of the complex on heating between 150 and 500°C is accompanied by a decrease in basal *d* values. Most of the interlayer humic macromolecules are highly resistant to alkali extraction and are, thus, humin-type materials. Therefore, the formation of humic substance interlayers in 2:1 expansible-layer silicates, through polymerization of phenol monomers and associated reactions in soils and sediments, deserves close attention.

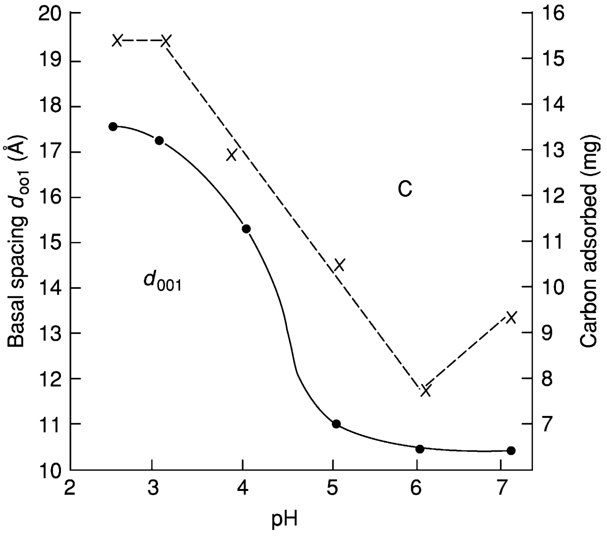


Figure 2 Effect of pH on basal spacing (d_{001}) and adsorption of fulvic acid. Reproduced with permission from Schnitzer M and Kodama H (1966) Montmorillonite: effect of pH on its adsorption of a soil humic compound. *Science* 153: 70–71.

Influence of Organic Matter on the Formation and Surface Chemistry of Metal Oxides

Metal oxides play a significant role in influencing soil behavior. They may exist as crystalline minerals, short-range ordered minerals, or noncrystalline precipitates which form surface coatings on clay minerals and humic substances. Organic components exert an important influence on the formation and surface properties of metal oxides.

Aluminum Oxides

The influence of a particular organic acid is generally related to the stability constant of the complex that the acid forms with Al (Table 1). Therefore, *p*-hydroxybenzoic acid, which forms the least stable complex with Al, does not inhibit the crystallization of Al hydroxides, whereas aspartic, tannic, malic, and citric acids increasingly retard crystallization. Besides the stability constant of the complex, the concentration of the organic acid is important. At certain low concentrations, the presence of some organic acids actually promotes the crystallization of particular $Al(OH)_3$ polymorphs by catalysis, but above a critical concentration, it disrupts crystallization through complexation with Al.

Humic substances also influence the transformation of Al by hampering the formation of more crystalline phases and promoting the formation of microcrystalline boehmite. FA and HA resemble aliphatic acids such as citric and malic acids, in that they contain COOH and aliphatic OH groups. They also resemble quercetin and tannic acid, because they contain phenolic

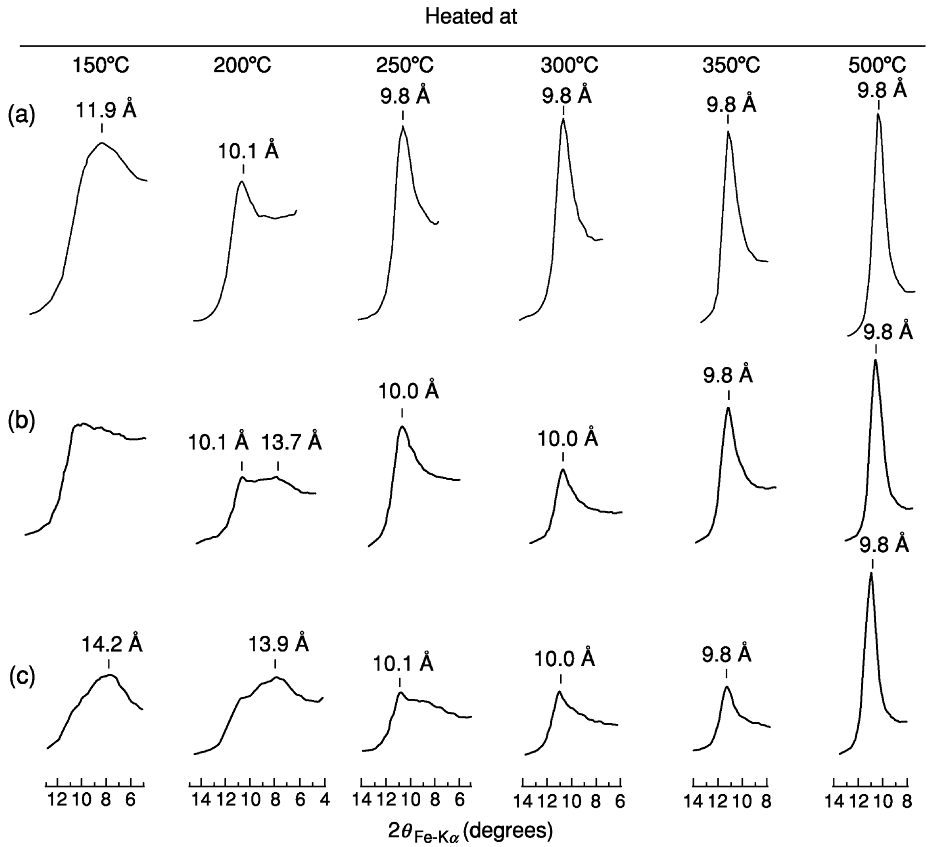


Figure 3 X-ray diffractograms of Ca-nontronite, influenced by deposition of humic macromolecules in the interlayers through polymerization of hydroquinone and the associated reactions. Diffractograms were recorded on a Philips X-ray diffractometer using Fe-Kα radiation with an Mn filter. (a) Ca-nontronite; humic macromolecule–Ca–nontronite complex before (b) and after (c) NaOH extraction. Reproduced with permission from Wang MC and Huang PM (1986) Humic macromolecular interlayering in nontronite through interaction with phenol monomers. *Nature* (London) 323: 529–531.

Table 1 Stability constants of the complexes formed between Al and five organic acids at 25°C

	Stability constants of the complexes	
Organic acids	*logK_1*	*logK_2*
p-hydroxybenzoic acid	1.66	–
Aspartic acid	2.60	–
Tannic acid	3.78	–
Malic acid	5.14	8.52
Citric acid	7.37	13.90

Reproduced with permission from Kwong KF, Ng Kee, and Huang PM (1979) The relative influence of low molecular weight, complexing organic acids on the hydrolysis and precipitation of aluminum. *Soil Science* 128: 337–342.

hydroxyl and ketonic C=O groups. Through these functional groups, FA and HA form stable complexes and inhibit the crystallization of Al hydroxides.

Organic components have an enormous impact on the surface properties of Al transformation products. For example, the presence of organic acids during aging of Al hydroxide gel for 40 days increases the specific surface of the products up to 30-fold over that of the control and higher organic acid concentrations result in higher specific surface. The surface charge characteristics of the products also dramatically change. The intermediate transformation products of Al, which include soluble mononuclear and polynuclear Al species and colloidal short-range ordered Al hydroxides, are the most reactive Al species in influencing physical, chemical, and biological processes of soil and related environments.

Iron Oxides

Soil organic matter plays a vital role in the formation of pedogenic Fe oxides (Figure 4). In soil environments where the amount of organic matter is low, for example in subsoils, the Fe supplied will form goethite and hematite depending on environmental factors. As organic matter content increases, more of the Fe will be complexed with organics, leading to decreases of Fe activity. The activity of Fe(III) ions

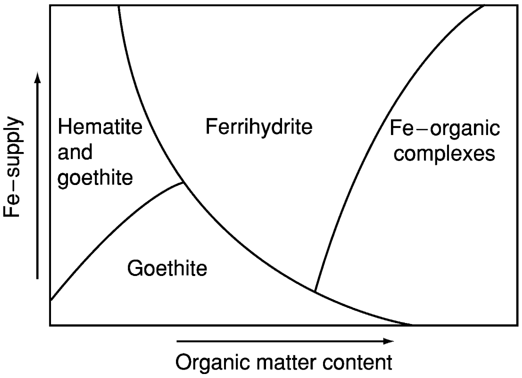


Figure 4 Schematic representation of the effect of organic matter content and rate of soluble Fe supply on the formation of various Fe forms in soils. Reproduced with permission from Schwertmann U, Kodama H, and Fisher WR (1986) Mutual interactions between organics and iron oxides. In: Huang PM and Schnitzer M (eds) *Interactions of Soil Minerals with Natural Organics and Microbes*, pp. 223–250. SSSA Special Publication 17. Madison, WI: Soil Science Society of America.

is so low that only the solubility product of goethite (10^{-41} to 10^{-42}), but not the solubility product of ferrihydrite (10^{-37} to 10^{-39}) is exceeded. Consequently, goethite but not ferrihydrite may form. Therefore, no hematite will form in an environment where the organic matter is high, since ferrihydrite is deemed to be a necessary precursor for hematite. This trend is generally observed in soils in temperate and cool regions as well as in wet depressions and surface soils of subtropical and tropical regions. At a higher content of organic matter, the rate of Fe supply is high; ferrihydrite will form and may survive for pedogenic times. If the content of organic matter is even higher, such as in O horizons or in peaty environments, all of the Fe may be in the form of Fe–organic complexes.

In an aqueous weathering environment, Fe oxides generally form via solution transformation. The oxidation products of Fe(II) solutions are thus important because it is in this valence state that Fe is commonly mobilized during weathering under the *E*h-pH region of natural soil environments. The kinetics of Fe(II) oxidation and the nature of Fe oxides are greatly influenced by a series of organic ligands. The rate constant of Fe(II) oxidation at the same ligand/Fe molar ratio generally decreases with the increase in the stability constant of the Fe(II)–ligand complex. The rate constant of Fe(II) oxidation also decreases with the increase in the perturbing ligand/Fe(II) molar ratio due to the increasing formation of the Fe(II)–ligand complex. The retardation of Fe(II) oxidation inhibits the nucleation of goethite and promotes the crystal growth of lepidocrocite. A further increase in the amount of the perturbing ligand distorts the structural order, resulting in the increasing inhibition of crystal growth of lepidocrocite and the formation of noncrystalline Fe oxides.

Surface properties of Fe oxides have a profound impact on microaggregate formation, water flux, nutrient and pollutant flux, and the ability of soils to promote plant growth, to maintain reasonable soil biotic habitat, and to respond to management, and resist degradation. The specific surface, charge characteristics, mean surface roughness, fractal dimension, and microporosity of Fe oxides depend on the nature and concentration of organic substances in the solution in which the Fe oxides are formed.

Role of Soil Minerals in Abiotic Catalysis of the Formation of Humic Substances

Polyphenol Pathway

Soil minerals play an important role in catalyzing the abiotic polymerization of phenolic compounds and the polycondensation of phenolic compounds and amino acids and the subsequent formation of humic substances. Layer silicates have the ability to catalyze the abiotic formation of humic substances. The promoting effect of 2:1 layer silicates is higher than 1:1 layer silicates. This is attributed to the higher specific surface and higher lattice imperfection of the former than the latter. The edges of kaolinite are virtually the only catalytic sites for the formation of phenol-derived humic substances. The edges of nontronite have a very important role as catalytic sites in the formation of hydroquinone-derived humic macromolecules. The Fe(III) in the octahedral sheet of nontronite also serves as a Lewis acid site to catalyze the oxidative polymerization of hydroquinone.

Aluminum, Fe, and Mn oxides, hydroxides, and oxyhydroxides also catalyze the oxidative polymerization of phenolic compounds. Among these minerals, Mn oxides are the most powerful catalysts in the transformation of phenolic compounds. Manganese oxides (birnessite, cryptomelane, and pyrolusite), which are common in soils, act as Lewis acids that accept electrons from phenolic compounds, leading to their formation of semiquinone and their oxidative polymerization and formation of humic substances. The catalytic power of metal oxides in affecting the C turnover and humic substances formation via abiotic processes in soil and related environments merits attention.

Poorly crystalline aluminosilicates are common in soils. Allophane has the ability to catalyze the polymerization of polyphenols. Proto-imogolite sol, which is present in soil solution, also plays a role in the abiotic formation of humic substances.

Primary minerals differ in their ability to catalyze the abiotic polymerization of hydroquinone. The sequence of the catalytic power of primary minerals is tephroite > actinolite > hornblende > fayalite > augite > biotite > muscovite ≃ albite ≃ orthoclase ≃ microcline ≃ quartz.

Maillard Reaction

The Maillard reaction is considered to be a major pathway in humification because of significant similarities between humic substances and melanoidins formed through this pathway, involving sugar–amino acid condensation. Birnessite (δ-MnO_2) significantly increases the extent of humification of the glucose–glycine system over the pH range of 6–8 (Figure 5). The chemical shifts of FA formed in the Maillard reaction system resemble those of natural humic substances. Further, light intensity of 168 μE $s^{-1}m^{-2}$ exerts an increase in the browning of the glucose–glycine–birnessite system compared with the same system kept in the dark. Equally important is that, even in complete darkness, birnessite catalyzes the Maillard reaction between glucose and glycine. Therefore, birnessite catalysis can occur in soil and sediment environments at any depth, but the presence of sunlight should accelerate the reaction.

Interactions of Soil Mineral Colloids with Microorganisms and Enzymes

Surface Interactions of Soil Minerals with Microorganisms

Soil is a habitat for myriads of microorganisms. The microbial biomass constitutes only a very small proportion (3%) of the total organic carbon in soil, but it is the most active and dynamic fraction of the living organic pool. Among the inorganic components of soil, mineral colloids are the most reactive because of their large specific surface and reactive charge characteristics. Being enriched in ions, water, and organic matter relative to the bulk soil, the surface of mineral colloids serves as a preferred habitat for soil microorganisms. Complex formation between mineral colloids and microorganisms can and does occur.

The surfaces of bacterial cells and crystalline clay minerals are both negatively charged. However, bacteria have a propensity for producing extracellular polysaccharides (EPS) which bind simultaneously

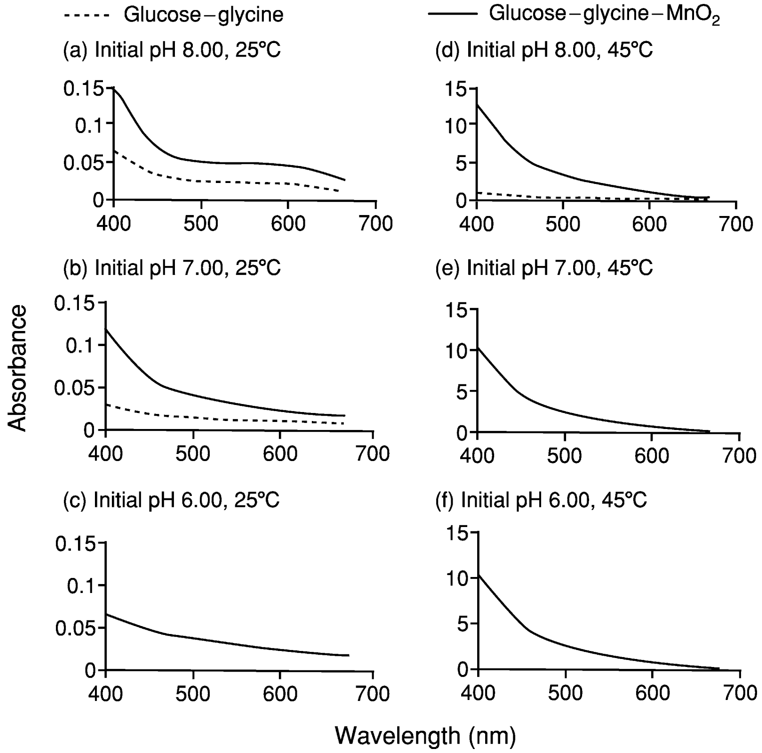


Figure 5 Absorbance versus wavelength plots in the Maillard reaction between glucose and glycine as influenced by birnessite catalysis. (a), (b), and (c): 30-day reaction period. (d), (e) and (f): 15-day reaction period. Reproduced with permission from Jokic A, Frenkel AI, Vairavamurthy MA, and Huang PM (2001) Birnessite catalysis of the Maillard reaction: its significance in natural humification. *Geophysical Research Letter* 28: 3899–3902.

to cell and clay surfaces through 'cation bridging,' involving polyvalent cations (Figure 6). Further, mineral colloids in soils are generally coated with hydroxy Al (or Fe) polymers (Figure 1). These coated minerals thus behave as a positively charged species or display amphoteric characteristics. Therefore, mineral colloids can strongly interact with negatively charged microbial cells in soil environments. This type of bonding, which is much stronger than cation bridging, is also expected to occur with Al and Fe oxides over the pH range of soils. The attachment of microorganisms to short-range minerals and the crystal edges of layer silicates through electrostatic interactions would also be predicted to occur when the soil pH falls below 6 because all of these surfaces would then be largely positively charged.

In the majority of cases, minerals in topsoils are partially covered with organic materials, especially humic substances (HS), which are microbially resistant and the most prevalent. The most common mode of mineral colloid–organic material–microorganism interactions may be represented by the following scheme:

$$[\text{Mineral colloid} - \text{HS}]^{-} \ldots\ \text{M}^{n+} \ldots {}^{-}[\text{EPS} - \text{B}]$$

In humic-rich calcareous Mollisols, Ca would be the predominant bridging cation. In Andisols, Oxisols, Ultisols, and the B horizon of spodosols, HS largely occur as complexes with Al and Fe, or their respective poorly crystalline oxides. In soils with little organic matter and in subsoils, mineral colloid–microorganism interactions are largely influenced by the mineralogical composition and pH of the system.

Mineral colloids can directly or indirectly influence the activity of microorganisms in their immediate vicinity. The effect of mineral colloids may be positive, negative, or sometimes so small as to escape detection. Mineral colloids have a stimulating effect on the activity of adhering bacteria by keeping the pH of microhabitats within the optimum physiological

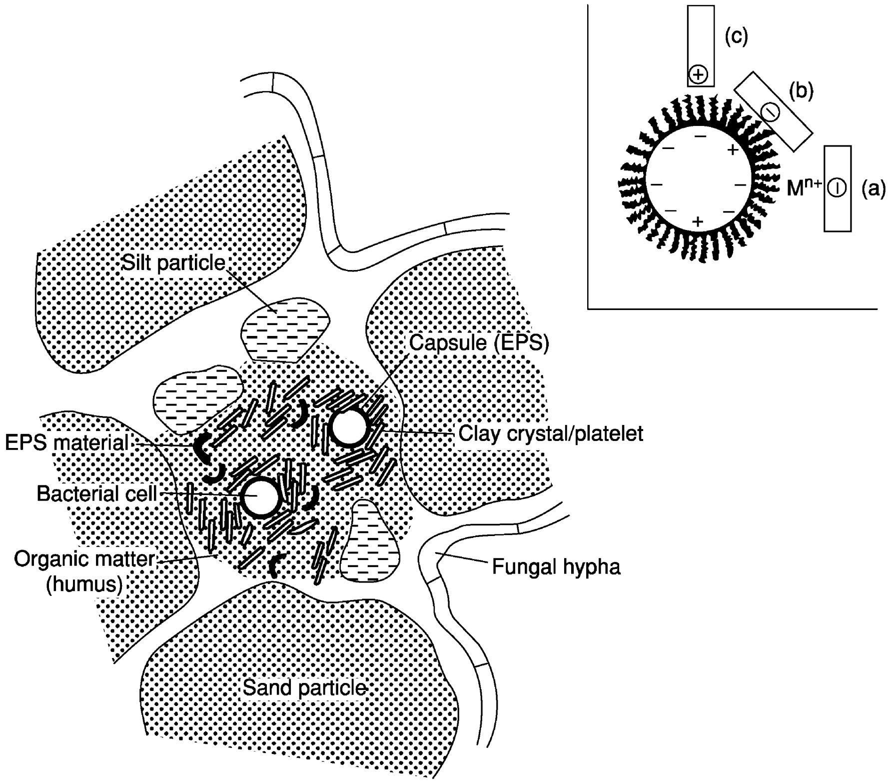


Figure 6 Diagram illustrating the interaction of bacteria and fungi with mineral particles in a soil aggregate. Bacterial cells with a coat of extracellular polysaccharides (EPS) are enveloped by clay particles. The pore space where clays and bacteria interact, bounded by silt- and sand-size particles, is relatively enriched in organic matter, including EPS residues. Fungal hyphae are attached to the outside surface of an aggregate. Inset: an enlarged view of a bacterial cell with its complement of EPS. At normal soil pH conditions, the cell has a net negative surface charge. Most clay particles adhere to the cell surface by bridging through polyvalent cations, represented by M^{n+} (a), although some may be attached directly by electrostatic interactions, either in a face-to-face (b), or edge-to-face (c) association. Reproduced with permission from Theng BKG and Orchard VA (1995) Interactions of clays with microorganisms and bacterial survival in soil: a physicochemical perspectives. In: Huang PM, Berthelin J, Bollag J-M, McGill WB, and Page AL (eds) *Environmental Impact of Soil Component Interactions*, vol. II. *Metals, Other Inorganics and Microbial Activities*, pp. 123–143. Boca Raton, FL: CRC/Lewis.

range for growth. Mineral colloids can also stimulate microbial activity by sorbing metabolites that would otherwise have an adverse effect on microbial growth. Further, the surfaces of mineral colloids can adsorb other toxic substances to microorganisms such as antibiotics and pesticides.

By forming an envelope around bacterial cells, mineral colloids may provide protection from extreme fluctuations in physicochemical environments, and thus enhance bacterial survival. The protective effect of mineral colloids, especially montmorillonite, is manifested in the ability of soil microorganisms to withstand exposure to hypertonic osmotic pressure, desiccation, and ultraviolet radiation. Mineral colloids can also act as a cementing agent of soil particles. Therefore, the addition of clay to soils, especially those of light texture, would modify the spatial arrangement of particles and especially the pore size distribution within soil aggregates. Such a modification of aggregate structure often benefits the bacterial population by increasing the proportion of pores of a certain size range ($<6\ \mu$m) from which bacteria could freely enter to colonize pores. However, bacterial predators – notably protozoa – are effectively barred from entering such pores due to steric hindrance.

The content and type of mineral colloids are influential in determining the balance between different microbial populations in soil. A well-known example is the failure of some fungi to thrive and spread in certain soils. This is largely attributed to the presence of montmorillonite in the 'suppressive' soils. Montmorillonite can serve as a proton sink and is thus able to promote the growth of acid-sensitive bacteria in these soils. This gives bacteria a selective advantage over fungi in competing for available nutrients, and fungal growth and proliferation are, thus, effectively suppressed.

Formation of Mineral Colloid– and Humic–Enzyme Complexes

Enzymes are rapidly sorbed at mineral colloids and humic substances in soil environments. Mineral colloids have a high affinity for enzymes, although this is not always synonymous with the retention of their enzymatic activity. Enzymes associated with soil humic substances retain their catalytic activity for long periods.

The mechanisms whereby enzymes are retained by mineral colloids have usually been assumed to include cation exchange reactions but not all observations have supported this view. Therefore, besides cation exchange reactions, adsorption of enzymes by mineral colloids may proceed through ionic, covalent, hydrophobic, and hydrogen bonding, and van der Waals forces. Although individual van der Waals forces are regarded as weak, retention may be cumulative if the enzyme is in close contact with the adsorbing surface. Recent studies show that enzymes such as tyrosinase have a very high affinity to the surfaces of hydrous manganese oxide birnessite. Analysis by atomic force microscopy (AFM) clearly shows the unfolding in an ellipsoidal shape of the tyrosinase protein after adsorption and immobilization on the mineral surfaces, especially the frayed edges of the phyllomanganate layer. AFM pictures along with X-ray diffraction and infrared absorption data of the tyrosinase protein complex show the coating of the mineral surfaces by the protein molecules as well as the changes in protein conformation.

Enzymes adsorbed or immobilized on mineral colloids operate in a structurally restricted system. The behavior of such enzymes toward their respective substrate is, thus, different from that occurring in a homogeneous, aqueous solution. When enzymes are adsorbed on mineral colloids, changes in the tertiary structure of the enzymes and their active sites decrease the activity or eliminate it altogether. However, there are notable exceptions to the 'adsorption-decline in activity' rule. Various supports show different enzyme immobilization capabilities (Table 2). There is considerable variation in the retained activities of the enzymes immobilized. The residual specific activities of laccase and peroxidase immobilized on all supports are high. Further, laccase immobilized on montmorillonite shows specific activity higher than that of the free enzymes. This is attributed to steric modification of the immobilized enzymes.

Many mechanisms have been proposed to account for the stability of enzymes which are complexed with soil humic substances. These include ion exchange, H-bonding, entrapment within three-dimensional micelles, lipophilic reactions, and covalent bonding during organic matter genesis. Functional groups of enzymes implicated in covalent bonding to humic polymers include terminal and basic amino, carboxyl, sulfhydryl, phenolic, and imidazole groups. All these may be involved in the stabilization process provided they do not form part of the active sites of the enzymes and not crucial to the retention of its tertiary structure. Enzyme–humic complexes can also be attached to mineral colloids and in some instances this further enhances enzyme stability.

Soil Mineral Weathering Transformations as Influenced by Microbial Activity

Mineral Weathering

Mineral weathering reactions can be enhanced by microbial activity by a factor as high as 10^6.

Table 2 Immobilization of a laccase (from *Trametes versicolor*) and a peroxidase (from horseradish) on different supports

Enzyme and support	*Protein adsorbed*[a] *(mg/%)*	*Enzymatic activity*		
		Units adsorbed[b]	*Specific activity*[c]	*Residual specific activity*[d] *(%)*
Laccase				
Glass beads	0.452/56	28.8	63.7	236
Montmorillonite	0.622/71	19.8	31.8	118
Kaolinite	0.566/64	13.1	23.1	85.5
Soil	0.644/73	15.7	24.4	90.4
Peroxidase				
Glass beads	0.092/17	8.4	91.6	93.8
Montmorillonite	0.224/43	23	102.8	105.2
Kaolinite	0.120/23	9.5	78.9	80.7
Soil	0.162/31	15	92.6	94.8

[a] Difference between proteins initially added to 200 mg of support (0.88 mg laccase and 0.52 mg of peroxidase) and those recovered in the supernatant and washings.

[b] Expressed as μmol O_2 consumed min^{-1} for laccase and μmol guaiacol transformed min^{-1} for peroxidase.

[c] Units adsorbed/protein adsorbed.

[d] Calculated as a percentage of the specific activity (sa) of the free enzyme (laccase, sa $= 27\ \mu\text{mol}^{-1}\ \text{min}^{-1}\ \text{mg}^{-1}$; peroxidase, sa $= 97.7\ \mu\text{mol min}^{-1}\ \text{mg}^{-1}$).

Reproduced with permission from Glantreda L and Bollag J-M (1994) Effects of soils on the behaviour of immobilized enzymes. *Soil Science Society of America Journal* 58: 1672–1681.

Microorganisms can dissolve minerals by direct and indirect action under aerobic and anaerobic conditions. The modes of attack of minerals by microorganisms include: (1) direct enzymatic oxidation or reduction of a reduced or oxidized mineral components; (2) indirect attack with a metabolically produced redox agent or inorganic and organic acids; (3) indirect attack by metabolically produced alkali, usually in the form of ammonia; (4) indirect attack with a metabolically produced ligand that forms a highly soluble product with a mineral component; and (5) indirect attack by biopolymers. The mode of microbial attack of a mineral may involve a combination of some of these mechanisms.

Uptake and concentration of elements by absorption and/or adsorption were observed with bacteria and fungi for P, Ca, Fe, and K. Fungi can adsorb K from solution, shift K equilibrium, and transfer micas to vermiculites. Such process can also occur for major and trace elements. When oxidized metal compounds such as Fe(III), Mn(IV), and As(V) act as electron acceptors, anaerobic respiration becomes an example of direct dissolving action under anaerobic conditions. Oxidation of sulfur entities of metal sulfides to obtain energy is an example of direct dissolving action under aerobic conditions. Volatilization of metals and metalloids or biomethylated metal and metalloid compounds from the soil into the atmosphere can be a mechanism of detoxification for toxic elements such as Hg, As, and Se.

Fine-Grained Mineral Formation

The physical and chemical characteristics of bacteria, such as their large surface area-to-volume ratio, serve to increase the metal-binding capacity of their charged surfaces, leading to precipitation and formation of mineral phases on their cell walls or surface polymers. The mechanisms by which bacteria initiate the formation of minerals in bulk solution vary widely between species. There may be a combination of biochemical and surface-mediated reactions during the process. Bacteria surface layers may passively adsorb and indirectly serve as a nucleation template. Bacteria can also more directly initiate mineral precipitation by producing reactive compounds, such as enzymes, metallothioneins, and siderophores, which bind metals and metalloids and catalyze their transformations. In addition, bacteria can instigate the spontaneous precipitation of metals and metalloids by altering the chemistry of their microenvironments.

An example of microbially mediated fine-grained mineral development is the formation of Mn oxides. Microbial oxidation of Mn(II) is a major process that can produce Mn oxide coatings on soil particles 10^5 times faster than abiotic oxidation. Manganese oxides are highly reactive minerals and can restrict the mobility of metals in soil and related environments. Biogenic Mn oxides have significantly larger specific surface and higher Pb adsorption capacity than abiotically precipitated Mn oxides. Bioformation of minerals thus merits attention in the remediation of metal contamination in the environment.

Impacts of Mineral–Organic Matter–Microbe Interactions on Soils in the Environment

Soil Structural Stability

The processes by which soil organic matter strengthens the bonds between soil mineral particles are complicated. Root exudation and microbial action produce organic compounds with a range of composition and molecular weights. These compounds interact with the mineral particles which also vary in size, shape, crystallinity, and electric charge. Interactions between soil mineral particles, organic substances, and microorganisms can occur at many different size scales, since the materials have a large size range in soils (Figure 7). The mechanisms for stabilization of soil structure vary with aggregate size.

The adsorption of organic molecules such as microbially derived polysaccharides and other unaltered

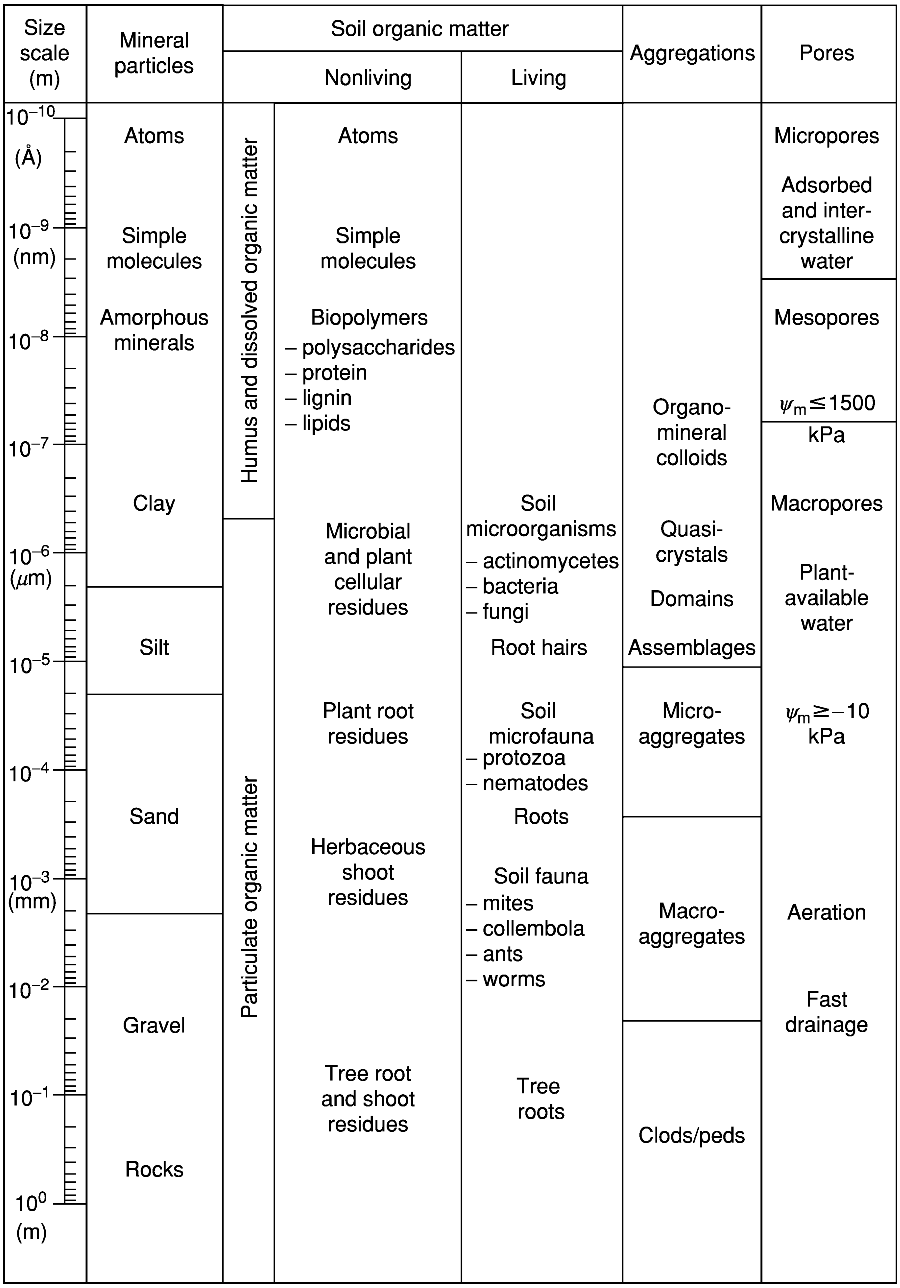


Figure 7 Size scales associated with soil mineral particles, organic components, pores, and aggregations of mineral and organic components. The definitions of pore size have used those developed by International Union of Pure and Applied Chemistry (IUPAC) (micropores <2 nm, mesopores 2–50 nm and macropores >50 nm). Alternatively the pore sizes corresponding to the lower ($\psi_m = -1500$ kPa) and upper ($\psi_m = -100$ kPa) limits of water availability to plants may be used to define the boundaries between the different classes of pore size. ψ_m is soil water metric potential. Reproduced with permission from Baldock JA (2002) Interactions of organic materials and microorganisms with minerals in the stabilization of structure. In: Huang PM, Bollag J-M, and Senesi N (eds) *Interactions Between Soil Particles and Microorganisms. Impact on the Terrestrial Ecosystem*, pp. 85–131. IUPAC Series on Analytical and Physical Chemistry of Environmental Systems. Chichester, UK: John Wiley.

and altered biomolecules on mineral surfaces can enhance the stability of individual clay microstructures. Adsorption is also important in the binding-together of clay microstructures and silt particles into small microaggregates with 2–50 μm diameters and densities $>20\,\mathrm{Mg\,m^{-3}}$. The high stability of these small microaggregates is demonstrated by their resistance to ultrasonification.

Mechanisms of stabilization of soil structure can operate over larger distances to bind microaggregates together to form macroaggregates (Figure 7). In view of the distances involved, the stabilization of macroaggregates is related to the presence of nonliving particulate organic matter (POM) capable of spanning distances $>100\,\mu$m or the existence of a network of fungal hyphae and plant roots that physically enmeshes microaggregates. The death of roots and hyphae growing within and through microaggregates results in the formation of biochemical binding agents capable of stabilizing the structure of macroaggregates.

Biochemical Cycling of C, N, P, and S in Soil

Chemically stabilized organic C in soils is slowly degraded to inorganic forms on time scales from centuries to millennia. Soil minerals are known to play a stabilizing role in soil organic matter. The interaction of Al and Fe with organic matter is of primary importance in the determination of the content of organic matter in tropical and temperate soils. The Al and Fe that complex and stabilize humus against microbial decomposition are released from soil minerals during soil formation and present as polymeric hydroxy cations. Their supply rates apparently control the content of soil organic matter to a great extent, as indicated by the relationship between pyrophosphate-extractable C and pyrophosphate-extractable Al and Fe.

The data from the radiocarbon analyses to explore interactions between soil mineralogy and soil organic C in volcanic soil environments show that the abundance of noncrystalline minerals accounts for >40% of the variation in organic C content across all the mineral horizons, substrate age, and soil orders (excluding the O and A horizons dominated by fast-cycling plant litter). Noncrystalline minerals also strongly influence turnover of organic matter.

Generally, more than 95% of the N and S and between 20 and 90% of the P in surface soils are present in soil organic matter. The close relationship between the organic forms of C, N, P, and S is well-established. The turnover of organic C is closely associated with the dynamics of N, P, and S in soils. The release of the stored organic nutrients depends on the mean residence time of soil organic matter associated with the mineral phase. The mean residence time of soil organic matter varies widely with the type of organomineral associations and the spatial location within the aggregate structure of soil (Table 3). Pools I and II are generally related to plant fragments, divided into easily available cell constituents and lignocelluloses at various degrees of degradation, respectively. Pool III consists of soil biomass and readily available organic matter within aggregates. On average, pools I–III account for about 20–30% of the total C in soil organic matter. These pools must be renewed continuously by fresh plant residues to maintain a relatively constant nutrient level and release by mineralization. The balance between decay and renewal processes, which is sensitive to management, controls the availability of N, P, and S. Pool IV is physically protected soil organic matter and is also affected by cultivation, because physical disturbance such as ploughing destroys macroaggregates and large microaggregates. Pool V is chemically stabilized soil organic matter which has the longest mean residence time. This pool represents 50–70% of the total C in soil organic matter and is seldom affected by management practices. Noncrystalline minerals have the ability to stabilize soil organic matter and thus reduce the turnover of C, N, P, and S.

Transformations of Organic Pollutants

Organic pollutants can be degraded through biotic and abiotic catalysis. Enzymes are biotic catalysts and mineral colloids are abiotic catalysts. It is frequently difficult to determine whether an organic pollutant is transformed abiotically or biotically. In many cases, significant abiotic and biotic catalytic reactions take place simultaneously.

The degradation of organic pollutants by extracellular enzymes is well-documented. Malathion, for instance, is hydrolyzed to a monocarboxylic acid (via loss of an ethyl group) by a carboxyl esterase isolated from soil. Enzymes commonly occurring in soil, such as esterases, amidases, phosphatases, and proteases, catalyze the hydrolysis of the respective chemical bonds in xenobiotic molecules.

Most organic chemicals, including xenobiotics, exhibit a strong affinity to humic substances. However, transformation of xenobiotics in terrestrial systems is greatly influenced by mineral components of soil. Mineral colloids, which are abundant in soil and have large specific surface and relatively high-charge density, contribute to the overall xenobiotic transformation at least as much as does the organic matter.

Soil organic matter and especially humic substances exert catalytic or inhibiting effects on abiotic hydrolysis of organic pollutants. Humic and fulvic acids have the ability to enhance the acid hydrolysis

Table 3 Comparison of estimated mean residence times of soil organic matter (SOM) in soil physical fractions

	Mean residence time (years)			
Pool	*Jenkinson and Rayner (1977)*[a]	*Parton et al. (1987)*[b]	*Buyanovsky et al. (1994)*[c]	Carter (1996)[d]
I	Decomposable plant material, 0.24	Metabolic plant residues, 0.1–1	Vegetative fragments 2–0.2 mm, 0.5–1	Litter, 1–3
II	Resistant plant material, 3.33	Structural plant residues, 1–5	Vegetative fragments 0.05–0.025 mm, 1–3	Free particulate organic matter (light fraction), 1–15
III	Soil biomass, 2.44	Active SOM pool, 1–5	OM in aggregates 2–1 mm, 1–4	Microbial biomass, 0.1–0.4
IV	Physically stabilized OM, 72	Slow SOM pool, 25–50	OM in aggregates 1–0.1 mm, 2–10	Intermicroaggregate OM[e], 5–50
V	Chemically stabilized OM, 2857	Passive SOM pool, 1000–1500	OM in fine silt, ~400 OM in fine clay, ~1000	Intramicroaggregate OM[f] physically sequestered, 50–1000 Chemically sequestered, 1000–3000

[a] Jenkinson DS and Rayner JH (1977) The turnover of soil organic matter in some of the Rothamsted classical experiments. *Soil Science* 123: 298–305.

[b] Parton WJ, Schimel DS, Cole CV, and Ojima DS (1987) Analysis of factors controlling soil organic matter levels in Great Plains grasslands. *Soil Science Society of America Journal* 51: 1173–1179.

[c] Buyanovsky GA, Aslam M, and Wagner GH (1994) Carbon turnover in soil physical fractions. *Soil Science Society of America Journal* 58: 1167–1173.

[d] Carter MR (1996) Analysis of soil organic matter storage in agroecosystems. In: Carter MR and Stewart BA (eds) *Structure and Organic Matter Storage in Agricultural Soils*, pp. 3–11. Advances in Soil Science. Boca Raton, FL: CRC Press.

[e] Organic matter stored within macroaggregates but external to microaggregates; includes coarse occluded particulate organic matter and microbial organic matter.

[f] Organic matter stored within microaggregates; includes fine occluded particulate organic matter and microbial-derived organic matter.

Reproduced with permission from Guggenberger G and Haider KM (2002) Effect of mineral colloids on biogeochemical cycling of C, N, P and S in soil. In: Huang PM, Bollag J-M, and Senesi N (eds) *Interactions between Soil Particles and Microorganisms. Impact on the Terrestrial Ecosystem*, pp. 267–322. IUPAC Series on Analytical and Physical Chemistry of Environmental Systems, vol. 8. Chichester, UK: John Wiley.

of phenoxy acetic acids and esters and chloro-s-triazines and to retard the alkaline hydrolysis of the *n*-octyl ester of 2,4-D. The mechanisms proposed to explain the catalytic effects of humic and fulvic acids on dechlorohydroxylation of the chloro-s-triazines, simazine, atrazine, and propazine, consist of the interaction through H-bonding between the surface carboxyl groups of humic and fulvic acids and the heterocyclic nitrogen atoms of the triazine ring.

Soil minerals play an important role in catalyzing the abiotic transformation of organic compounds. Their ability to catalyze the transformation substantially varies with their structural configuration and surface chemistry and the structure and functionality of the organic matter involved. Manganese oxides are most powerful soil minerals in catalyzing abiotic oxidation of phenolic compounds. Organic compounds such as 2,4-D and ethyl ether can be degraded by the catalysis of birnessite (δ-MnO_2). These organic compounds can be adsorbed on birnessite and rapidly oxidized, both producing CO_2 as a major product (Figure 8), but by somewhat different mechanisms. Therefore, the solid degradation of organochlorine herbicides can occur by catalysis of birnessite, which is common in the environment.

Mineral surfaces may also detoxify adsorbed organic pollutants by catalysis through their ability to behave as Brønsted acids and donate protons or to act as Lewis acids and accept electron pairs. Brønsted acidity derives essentially from the dissociation of water molecules coordinated to surface-bound cations. Therefore, this acidity is strongly influenced by the hydration status and polarizing power of surface-bound and structural cations on mineral colloids. Certain organic pollutants such as organophosphate and s-triazine pesticides can be degraded by catalysis of mineral colloids through their surface acidity. This type of reaction deserves attention in a wide range of other organic compounds in soil environments. Besides the Brønsted acidity, the Lewis acidity of metals such as Al and Fe exposed at the edges of minerals is important in mineral-catalyzed hydrolysis reactions, at least for those that are hydroxyl ion-catalyzed. The ability of a metal ion to catalyze the hydrolysis varies with its ability to complex with reactant molecules and shift electron density and conformation in ways favorable to the reaction.

The degradation of organic pollutants may be considerably reduced when they are retained by soil mineral and humic colloids. The major reason for reduced biodegradation rates is the diminished bioavailability of chemicals involved in binding processes. The availability of sorbed xenobiotics to microorganisms varies with the chemical properties of the pollutant, the

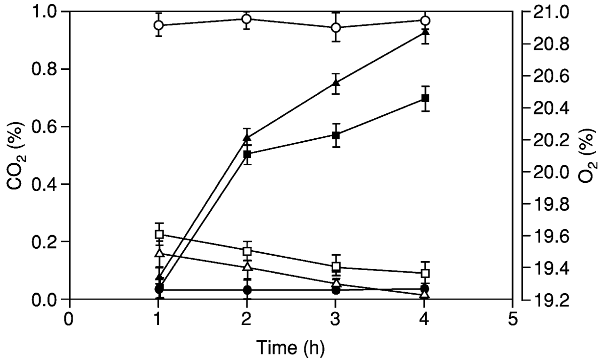


Figure 8 Gas chromatographic measurements of CO_2 and oxygen in the headspace of reaction vials and 2,4-D at 1, 2, 3, and 4 h, respectively. The values indicate the mole percentage of CO_2 and oxygen in the headspace gas. Filled circles, CO_2 from air control; filled squares, CO_2 from ethyl ether control; filled triangles, CO_2 from 0.3 μmol 2,4-D plus ethyl ether sample; open circles, oxygen from air control; open squares, oxygen from the ethyl ether control; open triangles; oxygen from 0.3 μmol 2,4-D plus ethyl ether. The error bars indicate the standard error for two measurements. Reproduced with permission from Cheney MA, Sposito G, McGrath AE, and Criddle RS (1996) Abiotic degradation of 2,4-D (dichlorophenoxyacetic acid) on synthetic birnessite: calorespirometer method. *Colloids and Surfaces. A. Physicochemical and Engineering Aspects* 107: 131–140.

nature of the sorbent, the mechanism of sorption, and the properties of the organisms. The impact of structural configuration and surface properties of soil mineral and organic components on the activity of microorganisms and their ability to degrade organic pollutants with different structure and functionality merit increasing attention for years to come.

Transformations of Metals

The transformation of metals in the environment is influenced by soil mineral–organic matter–microbe interactions. The impacts of these interactions on metal transformation are especially important in the soil rhizosphere because of root exudation. This leads to colonization by different populations of bacteria, fungi, protozoa, and nematodes. Plant–microbe interactions result in intense biological processes in the rhizosphere. These interactions, in turn, affect physicochemical reactions in the rhizosphere. Physicochemical properties that can be different in the rhizosphere include acidity, concentration of complexing biomolecules, redox potential, ionic strength, moisture, and nutrient status. The total rhizosphere environment is governed by an interacting trinity of the soil, the plant, and the organisms associated with the root. Therefore, reactions in the soil rhizosphere, which is the bottleneck of metal contamination of the terrestrial food chain, can only be satisfactorily integrated based on physical, chemical, and biological interactive processes.

A series of complexation reactions in soils affect metal transformation in the rhizosphere. Complexation reactions of metals with organic ligands are significant in determining the chemical behavior, availability, and toxicity of metals in the rhizosphere. The plant and prolific microbial activity results in increased amounts of organic ligands at the soil–root interface. Therefore, a larger fraction of the metal contaminant is in a complexed and usually soluble form in the rhizosphere soil than in the bulk soil.

The transformation of metals is significantly influenced by adsorption–desorption reactions in soil environments. These reactions are affected by biochemical and biological processes, especially in the rhizosphere. There is increasing evidence that organic ligands modify the adsorption–desorption reactions of metals in soils. For most of the trace metals, direct precipitation from solution through homogeneous nucleation appears to be less likely than adsorption–desorption by virtue of the low concentrations of these metals in soil solutions in well-aerated dryland soils. In aerobic soils, although precipitation of trace metals through homogeneous nucleation is not likely, heterogeneous nucleation may play a significant role in metal transformation because of the presence of mineral, organic, and microbial surfaces that catalyze the nucleation set of crystallization. The energy barrier to nucleation is reduced or removed by surfaces. The catalytic processes reduce the extent of supersaturation necessary for precipitation to occur.

Besides physicochemical reactions, metals have easy access to microbial surfaces through diffusion. All microorganisms contain biopolymers, such as proteins, nucleic acids, and polysaccharides, which provide sites where metal ions can bind. These binding sites include negatively charged groups such as carboxylate, thiolate, phosphate, and groups such as amines which coordinate to the metal center through lone pairs of electrons. Because of the ability of these biopolymers to bind metals, large concentrations of metals are frequently associated, not only with living microbial biomass, but also with dead cells. Many metals bind with various degrees of tenacity to the largely anionic outer-surface layers of microbial cells. Some metals are bound by cell walls to a greater extent than by clay minerals (Table 4), indicating that bacterial cell walls and membranes may act as foci for accumulation of metals in soils. Surface metal concentrations frequently exceed the stoichiometry expected per reactive chemical sites within cell walls. The sorption of metals can be so great that precipitates can be formed and distinct metallic minerals are eventually formed by biomineralization (the

Table 4 Metals bound by native *Bacillus subtilis* walls, *Escherichia coli* envelopes, kaolinite, and smectite

	Amount of metal bound (μ*mol* g^{-1}) *(oven-day weight)*[a]			
Metal	*Walls*	*Envelopes*	*Kaolinite*	*Smectite*
Ag	423 ± 15	176 ± 3	0.46 ± 0.02	43 ± 0.3
Cu	530 ± 13	172 ± 9	5 ± 0.03	197 ± 4
Ni	654 ± 25	190 ± 3	4 ± 0.2	173 ± 10
Cd	683 ± 19	221 ± 6	6 ± 0.2	1 ± 0.02
Pb	543 ± 11	254 ± 5	3 ± 0.2	118 ± 6
Zn	973 ± 13	529 ± 32	37 ± 1	65 ± 2
Cr	435 ± 37	102 ± 2	8 ± 0.5	39 ± 5

[a] The data represent the average of three to five determinations for each sample from duplicate experiments and the standard error.

Reproduced with permission from Walker SG, Flemming CA, Ferris FG, Beveridge TJ, and Bailey GW (1989) Physicochemical interaction of *Escherichia coli* cell envelopes and *Bacillus subtilis* cell walls with two days and ability of the composite to immobilize heavy metals from solution. *Applied and Environmental Microbiology* 55: 2976–2984.

formation of minerals by microorganisms). Biomineralization has global consequences in dynamics, toxicity, and fate of metal pollutants.

See also: **Microbial Processes:** Environmental Factors; Community Analysis; Kinetics; **Organic Matter:** Principles and Processes; Genesis and Formation

Further Reading

Cheney MA, Sposito G, McGrath AE, and Criddle RS (1996) Abiotic degradation of 2,4-D (dichlorophenoxyacetic acid) on synthetic birnessite: calorespirometer method. *Colloids and Surfaces. A. Physicochemical and Engineering Aspects* 107: 131–140.

Gianfreda L and Bollag J-M (1994) Effect of soils on the behavior of immobilized enzymes. *Soil Science Society of America Journal* 58: 1672–1681.

Huang PM, Bollag J-M, and Senesi N (eds) (2002) *Interactions between Soil Particles and Microorganisms. Impact on the Terrestrial Ecosystem*. IUPAC Series on Analytical and Physical Chemistry of Environmental Systems, vol. 8. Chichester, UK: John Wiley.

Huang PM, Wang MK, Kämpf N, and Schulze DG (2002) Aluminum hydroxides. In: Dixon JB and Schulze DG (eds) *Soil Mineralogy with Environmental Applications*, pp. 261–289. SSSA Book Series 7. Madison, WI: Soil Science Society of America.

Jastrow JD and Miller RM (1997) Soil aggregate stabilization and carbon sequestration: feedbacks through organomineral associations. In: Lal R, Kimble JM, Follett RF, and Stewart BA (eds) *Soil Processes and the Carbon Cycle*, pp. 207–223. Boca Raton, FL: CRC Press.

Jokic A, Frenkel AI, Vairavamurthy MA, and Huang PM (2001) Birnessite catalysis of the Maillard reaction: its significance in natural humification. *Geophysical Research Letter* 28: 3899–3902.

Naidja A, Liu C, and Huang PM (2002) Formation of protein–birnessite complex: XRD, FTIR, and AFM analysis. *Journal of Colloid and Interface Science* 251: 46–56.

Schnitzer M and Kodama H (1966) Montmorillonite: effect of pH on its adsorption of a soil humic compound. *Science* 153: 70–71.

Schwertmann U, Kodama H, and Fisher WR (1986) Iron oxides. In: Huang PM and Schnitzer M (eds) *Interactions of Soil Minerals with Natural Organics and Microbes*, pp. 223–250. SSSA Special Publication 17. Madison, WI: Soil Science Society of America.

Stotzky G (1986) Influence of soil mineral colloids on metabolic processes, growth, adhesion, and ecology of microbes and viruses. In: Huang PM and Schnitzer M (eds) *Interactions of Soil Minerals with Natural Organics and Microbes*, pp. 305–428. SSSA Special Publication 17. Madison, WI: Soil Science Society of America.

Tebo BM, Ghiorse WC, van Waasbergen LG, Siering PL, and Caspi RC (1997) Bacterially mediated mineral formation: insights into manganese oxidation from molecular genetic and biochemical studies. In: Banfield JF and Nealson KH (eds) *Geomicrobiology: Interactions between Microbes and Minerals, Reviews in Mineralogy*, vol. 35, pp. 225–266. Washington, DC: Mineralogical Society of America.

Theng BKG and Orchard VA (1995) Interactions of clays with microorganisms and bacterial survival in soil: a physicochemical perspective. In: Huang PM, Berthelin J, Bollag J-M, McGill WB, and Page AL (eds) *Environmental Impact of Soil Component Interactions*, vol. II, *Metals, Other Inorganics and Microbial Activities*, pp. 122–143. Boca Raton, FL: CRC/Lewis.

Torn MS, Trumbore SE, Chadwiçk OA, Vitousek PM, and Hendricks DM (1997) Mineral control of soil organic carbon storage and turnover. *Nature (London)* 389: 170–173.

Walker SG, Flemming CA, Ferris FG, Beveridge TJ, and Bailey GW (1989) Physicochemical interaction of *Escherichia coli* cell envelopes and *Bacillus subtilis* cell walls with two clays and ability of the composite to immobilize heavy metals from solution. *Applied and Environmental Microbiology* 55: 2976–2984.

Wang MC and Huang PM (1986) Humic macromolecular interlayering in nontronite through interaction with phenol monomers. *Nature (London)* 323: 529–531.

MINERALS, PRIMARY

P M Huang, University of Saskatchewan, Saskatoon, SK, Canada
M K Wang, National Taiwan University, Taipei, Taiwan

Introduction

Primary minerals are formed at elevated temperatures and inherited from igneous and metamorphic rocks, sometimes through a sedimentary cycle. In most soils, the sand and silt fractions consist largely of primary minerals. In the clay fraction of weakly weathered soils, primary minerals are present but are minor constituents of the clay fraction of most agricultural soils.

The most abundant primary minerals in soils are quartz and feldspars, just as they are the dominant rock-forming minerals in the Earth's crust. Quartz consists of a continuous framework of silica tetrahedra and is the main form of free silica occurring in soils. Feldspars are anhydrous, three-dimensional aluminosilicates containing varying amounts of Na, K, and Ca, and occasionally of other large cations such as Ba. Feldspars are found in virtually all sediments and soils in quantities that vary with the nature of the parent material and the stage of weathering. Weathering of feldspars is an important Earth-surface geochemical process. Feldspars are stores of the macronutrients K and Ca; they play a substantial role in overall K dynamics of soils.

Micas are 2:1 phyllosilicates with tightly held, nonhydrated, interlayer cations balancing a high layer charge. They are largely of primary origin, occur extensively, and serve as precursors for expansible 2:1 phyllosilicates, vermiculites, and smectites in soil environments. Micas are often present in soils as components of particles that have been only partially transformed to expansible 2:1 minerals through interstratification with the other minerals or the formation of mica cores surrounded by expanded zones. Micas are important reserves of soil K for growing plants. The K-supplying power of a soil depends to a considerable extent on the nature and amount of micas present and their dynamics of K release.

Olivines, pyroxenes, and amphiboles are important accessory primary minerals that occur in small but significant amounts. Olivines are olive-green nesosilicates in which Mg and Fe^{2+} are octahedrally coordinated by O atoms. Pyroxenes and amphiboles, which are inosilicates, are ferromagnesian minerals with single- and double-chain structures, respectively, of linked silica tetrahedra. The variety of isomorphous substitution possible in olivines, pyroxenes, and amphiboles and their relative ease of weathering make them excellent source minerals for Ca, Mg, and trace elements in soils. Furthermore, the ability of nesosilicates and inosilicates to catalyze the abiotic formation of humic substances in soils merits attention.

Quartz

Silica occurs in nature as seven distinct polymorphs: quartz, cristobalite, tridymite, coesite, stishovite, lechatelerite (silica glass), and opal. Disordered cristobalite commonly occurs in soils; tridymite is rare in soils and usually associated with siliceous volcanic rocks; and coesite, stishovite, and lechatelerite are rare polymorphic forms. Opal is a hydrated, amorphous silica and is not uncommon, but, of these minerals, quartz is the most abundant in soil environments.

Quartz is one of the last minerals to crystallize from magma. Therefore, it is formed under conditions closer to present Earth-surface conditions than minerals crystallized earlier. This contributes to its high stability. Compared with other silica polymorphs, dense packing of the crystal structure and high activation energy required to alter the Si-O-Si bond are major factors contributing to the high stability of quartz. Quartz is present in essentially all soils and often constitutes the major portion of all sand and silt fractions. It is a major component of the coarse clay fraction of many soils.

The silica tetrahedron of quartz is almost symmetrical and has an Si-O distance of 0.161nm. The Si-O bond is approximately equally ionic and covalent in nature. The structure of quartz can be visualized as a spiral network of silica tetrahedra about the *z*-axis (Figure 1). Each tetrahedron is repeated in the network by a rotation of 120° and a translation of one-third about the *z*-axis. The tetrahedra are linked to form a hexagonal structure in quartz.

The solubility of silica minerals decreases as a function of increasing packing density of the silica tetrahedra and long-range crystal order. Solubility of quartz at ambient temperatures and neutral pH is commonly 3–7 $\mathrm{mg\,Si\,l^{-1}}$, which is much lower than that of amorphous silica (50–60 $\mathrm{mg\,Si\,l^{-1}}$). Freshly ground quartz commonly shows abnormally high solubility levels (37 $\mathrm{mg\,Si\,l^{-1}}$). This has been attributed to the formation of a disrupted surface layer, which is believed

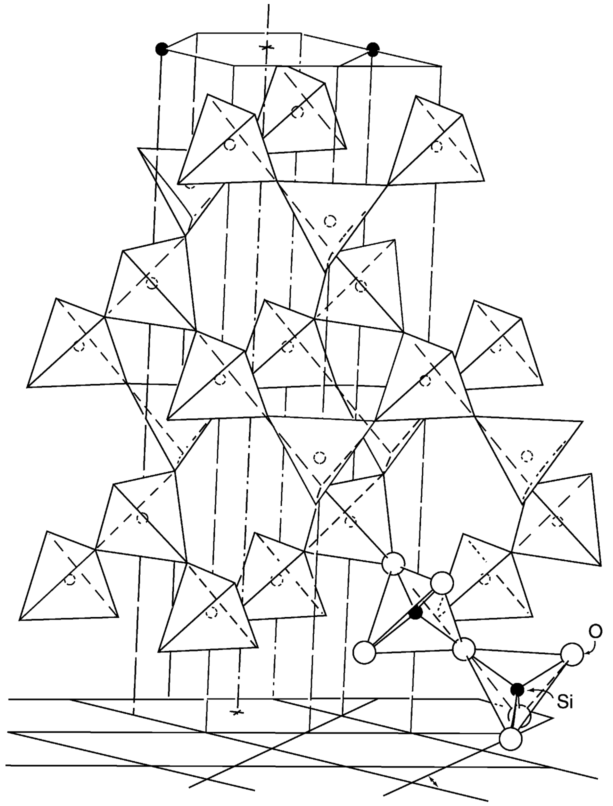


Figure 1 The structure of low quartz, SiO_2, shown as SiO_4 tetrahedra with a small silicon atom at the center of a group of four oxygens in tetrahedral coordination. Each tetrahedron shares a corner (oxygen) with an adjoining tetrahedron. For simplicity most of the tetrahedrons are shown without the circles representing oxygen. The screw axis symmetry (threefold) of the *c*-axis is shown by the dashed line, repeating each tetrahedron with a rotation of 120° and a translation of *c*/3. Two units of structure along *c* are shown. (Reproduced with permission from Berry LG and Mason B (1959) *Mineralogy*. San Francisco, CA: WH Freeman.)

to be amorphous or microcrystalline. Further, the solubility of quartz is a function of temperature, pH, and particle size. It increases with increasing temperature and is essentially constant between the pH values of 2 and 8.5, but increases rapidly above 9, due to the ionization of monosilicic acid:

$$Si(OH)_4 + OH^- = Si(OH)_3O^- + H_2O$$

The dissolution rate of quartz increases with the decrease in particle size. This is mainly attributed to an increase in specific surface of quartz as particle size decreases.

The presence of organic molecules such as amino acids greatly enhances the dissolution of silica, including quartz. The high rate of quartz dissolution in soils is attributable to leachates rich in organic molecules which react with quartz to form Si–organic molecular complexes in soil solution. The dissolution of quartz is greatest in the root zone, where production of biomolecules increases the dissolution of quartz by complexing with monosilicic acid.

Sesquioxides are a significant component of most soils and play a role in determining the concentration of dissolved silica in soils. They are responsible for much of the capacity of soils to sorb soluble silica, with the maximum adsorption occurring between pH values of 8 and 10. They act as a sink of dissolved silica and greatly increase the dissolution rate of silica. Therefore, finely divided quartz may not be stable at these pH values in the presence of excess sesquioxides.

Oxygen-isotope composition of minerals is a useful tool for determining the genesis and history of soil parent materials. The oxygen-isotope composition of quartz is dependent on the temperature of formation; delta values ($\delta^{18}O$) decrease as the temperature of formation increases. Delta values ($\delta^{18}O$) are defined as follows:

$$\delta^{18}O = \left[\frac{^{18}O/^{16}O\ \text{sample}}{^{18}O/^{16}O\ \text{SMOW}} - 1\right] \times 1000$$

where SMOW stands for standard mean ocean water. Quartz in igneous rocks has delta values in the order of 8–13‰; whereas quartz from metamorphic rocks has somewhat higher delta values, with a wider range that results from varying degrees of metamorphism. Oxygen-isotope ratios of quartz in shales range from 15 to 24‰; the ratios of quartz in sandstones range from 10 to 16‰. Silica minerals formed at ambient temperatures (chert, diatom, plant opal) have considerably higher delta values. The intermediate values for sedimentary deposits are attributed to the mixed origins of the sediments. Delta values of quartz from soils range from 9 to 30‰, which are generally similar to those of sedimentary rocks. Quartz is exceedingly resistant to oxygen-isotope exchange; therefore, the isotopic composition of quartz in soils, dusts, and sediments has been used as an indicator of provenance. The main use of isotopic composition of quartz in pedologic studies has been for identification of eolian additions to soils.

Feldspars

Feldspars make up an average of $595\,g\,kg^{-1}$ of igneous rock, $300\,g\,kg^{-1}$ of shale, and $115\,g\,kg^{-1}$ of sandstone. Many metamorphic rocks also contain large amounts of feldspars. Therefore, feldspars are present in virtually all sediments and soils in quantities that vary with the nature of parent material of soils and their degree of weathering. They play a substantial role in overall dynamics of macronutrients, K and Ca, in soils.

The structure of feldspars is a three-dimensional framework of linked SiO_4 and AlO_4 tetrahedra, with sufficient opening in the framework to accommodate K, Na, Ca, or Ba to maintain electroneutrality. In the building up of the framework, four-membered rings of tetrahedra are the basic units. Chains of four-membered rings of tetrahedra are parallel to the *a*-axis and cross-linked to adjacent parallel chains by shared O atoms (Figure 2). One zigzag chain is formed by superimposing four-membered rings of tetrahedra that share some of their vertices to form new, four-membered rings. The remaining vertices form cross-links between chains. The repeat periodicity along the chain is 0.84 nm in the ideal case. The range of the *a* axis of different feldspars approximately varies from 0.81 to 0.86 nm. These four-membered rings are linked together to form a honeycomb-type arrangement.

Feldspars represent limited solid solution between the three end members (K-feldspar, albite, and anorthite) with pure K, Na, and Ca, respectively. Alkali feldspars are the K- and Na-rich members of the group with a small amount of Ca. Plagioclases are members of this group of minerals, rich in Ca and/or Na, but with small amounts of K.

Alkali feldspars have a range of chemical composition and the end members may be written as $KAlSi_3O_8$ and $NaAlSi_3O_8$. The K-feldspar polymorphs, sanidine, orthoclase, microcline, and adularia have identical chemical compositions. Sanidine is a monoclinic alkali feldspar with small optic axial angle (2V) and commonly occurs in volcanic rocks. Orthoclase is a monoclinic alkali feldspar that has the larger optical axial angle and looks homogeneous; it does not show cross-hatched twinning. Microcline is triclinic, has the larger optic axial angle (2V), and exhibits the typical cross-hatched twinning (Figure 3a). The alkali feldspars, which can be either monoclinic or triclinic but have a special crystal habit and occur in low-temperature hydrothermal veins, are referred to as 'adularia.' Anorthoclase is a triclinic, Na-rich feldspar that shows a very fine cross-hatched twinning.

Plagioclase feldspars have a chemical composition between pure albite and pure anorthite. The classification of the plagioclase series is based on the mole fractions of the albite (A_b) and anorthite (A_n) components (Table 1).

The high-temperature series of mixed crystals is quite continuous between the end-member orthoclase

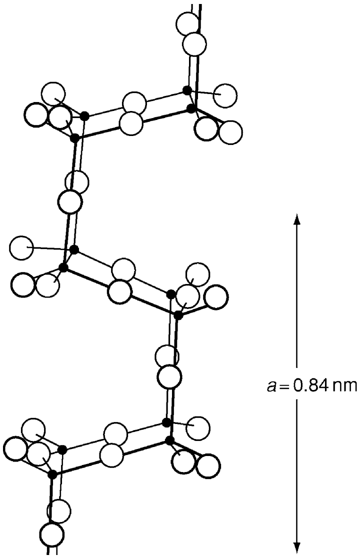


Figure 2 The essential structural feature of all feldspars projected on the (001) plane. The small black circles are Si and Al, the large circles, O. The O atoms projecting to right and left form the means of linking this chain to its neighbors. (Reproduced from Taylor WH (1965) Framework silicates: the feldspars. In: Bragg L and Claringbull GF (eds) *Crystal Structures of Minerals: The Crystalline State*, vol. IV. London, UK: G Bell.)

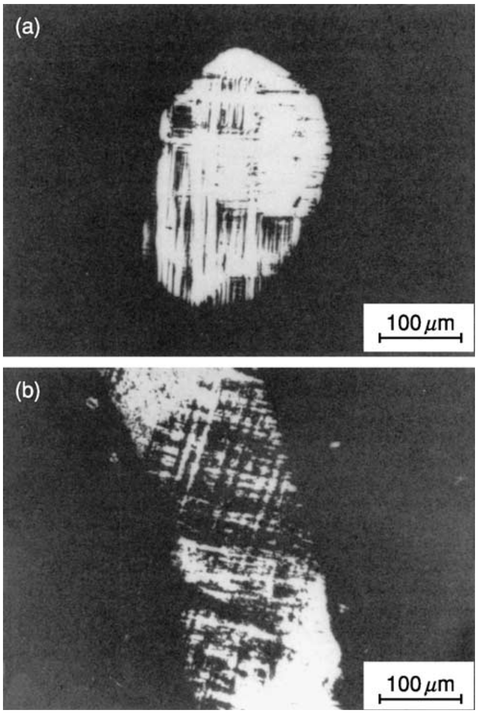


Figure 3 Photomicrographs of K-feldspar crystals from the Ap horizon of an Aridic Haploboroll (Orthic Brown Haverhill soil) in Saskatchewan, Canada: (a) a K-feldspar crystal with the cross-hatched pattern of maximum microcline of fine sand (50–250 μm) fraction; (b) a K-feldspar crystal, showing two phases of a coarse to very coarse sand (500–2000 μm) fraction. (Reproduced with permission from Somasiri S and Huang PM (1973) The nature of K-feldspars of selected soils in the Canadian prairies. *Soil Science Society of America Proceedings* 37: 461–464.)

Table 1 The classification of plagioclase feldspars

Plagioclase feldspars	*Mole fractions of albite (Ab) and anorthite (An)*
Albite	Ab_{100}–$Ab_{90}An_{10}$
Oligoclase	$Ab_{90}An_{10}$–$Ab_{70}An_{30}$
Andesine	$Ab_{70}An_{30}$–$Ab_{50}An_{50}$
Labradorite	$Ab_{50}An_{50}$–$Ab_{30}An_{70}$
Bytownite	$Ab_{30}An_{70}$–$Ab_{10}An_{90}$
Anorthite	$Ab_{10}An_{90}$–Ab_{100}

and albite and also between albite and anorthite. However, under low-temperature conditions, these phases undergo structural changes and tend to unmix. The unmixing results in the formation of a lamellar aggregate whose composition is similar to that of the end members. The lamellar aggregate, which is composed of a large amount of alkali feldspar and a subordinate amount of albite, is referred to as 'perthite.' Except for authigenic K-feldspar, which is formed through reconstitution and precipitation, most alkali feldspars usually contain varying amounts of Na in their structure and are thus more or less perthitic. The morphology of soil perthitic crystals that appear to contain a second phase, most likely albite, in a host crystal of K-feldspar is shown in Figure 3b.

Feldspars are present in nearly all soils, but their quantities vary with the intensity and capacity factors of weathering reactions. They are completely absent or present only in small quantities in strongly weathered soils, though the parent material contains considerable quantities of those minerals. However, some feldspars may be present in humid tropical soils that contain relatively fresh rock materials due to erosional and depositional processes. Feldspars are commonly present in the silt and sand fractions of young to moderately developed soils representing various soil parent materials and soil-forming conditions. Alkali feldspars are even present in the clay fraction of moderately weathered soils.

Persistence of feldspars in soils is related to the nature of the minerals, climate, topography, degree of leaching, complexation, redox potential, and certain solution ionic activities. According to the Goldich stability series, the increasing order of stability of feldspars is: anorthite < bytownite < labradorite < andesine < oligoclase < albite < K-feldspars. This order is related to the decrease in Al and Ca contents. Within the K-feldspar group, microcline is more stable than orthoclase. This is attributed to the difference in the volume occupied by the O atoms in the minerals. Compared with the monoclinic symmetry of orthoclase, the smaller volume is occupied by the O atoms of microcline by virtue of its triclinic symmetry.

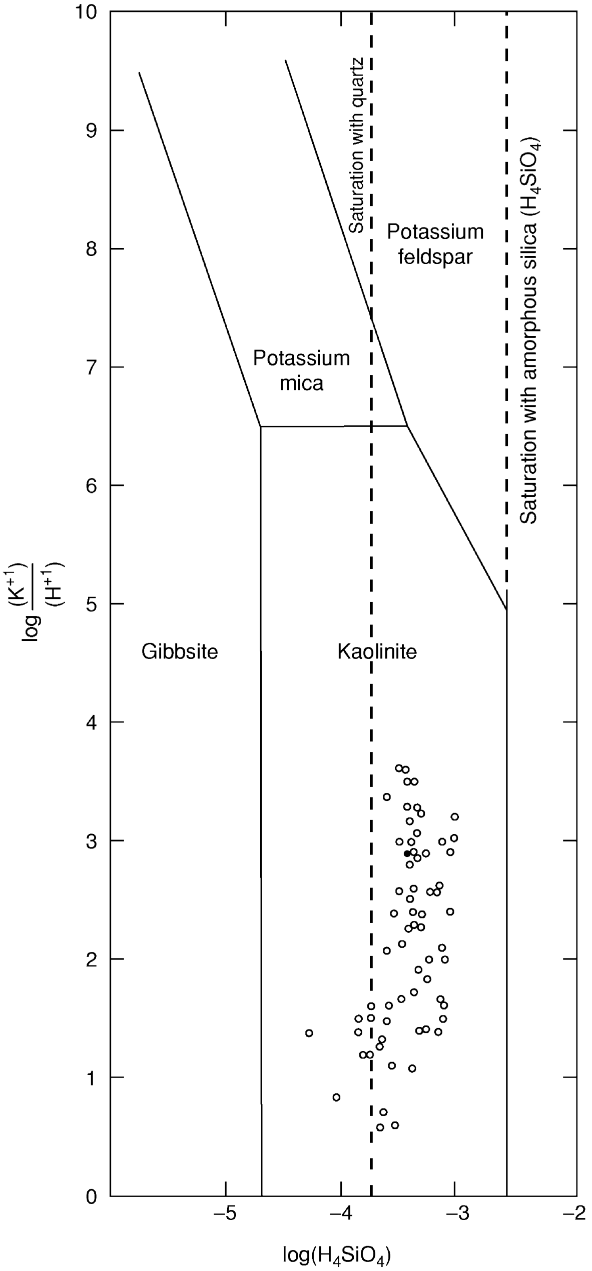


Figure 4 Stability relations of some phases in the system K_2O-Al_2O_3-SiO_2-H_2O at 25°C and 0.1 MPa pressure as functions of $(K^+)/(H^+)$ and (H_4SiO_4). Circles represent analyses of groundwaters. (Reproduced with permission from Feth JH, Robertson CE, and Polzer WL (1964) *Sources of Mineral Constituents in Water from Granitic Rocks, Sierra Nevada, California and Nevada*. US Geological Survey Water Supply Paper, 1535-I.)

Since establishment of the empirical stability series of Goldich, relative stabilities of feldspars at 25°C and 0.1 MPa have been thermodynamically shown in stability diagrams. The stability diagrams of K-feldspar and albite are shown in Figures 4 and 5, respectively. These stability diagrams are plotted as a function of the activity of H_4SiO_4 and the ratio of the

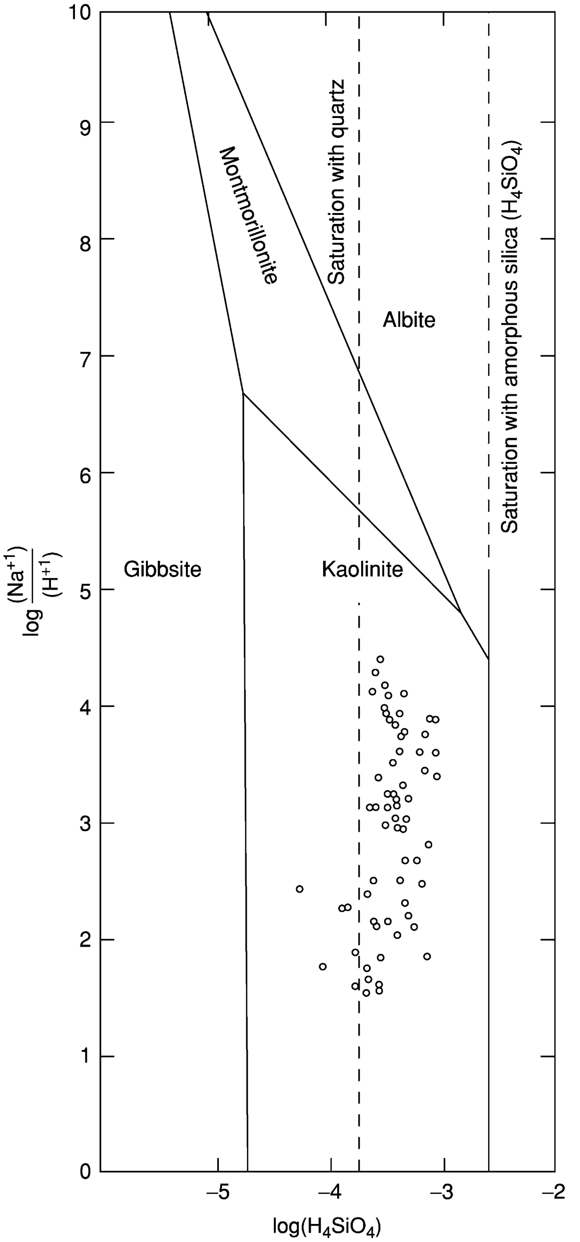


Figure 5 Stability relations of some phases in the system Na_2O-Al_2O_3-SiO_2-H_2O at 25°C and 0.1 MPa pressure as functions of $(Na^+)/(H^+)$ and (H_4SiO_4). Circles represent analyses of groundwaters. (Reproduced with permission from Feth JH, Robertson CE, and Polzer WL (1964) *Sources of Mineral Constituents in Water from Granitic Rocks, Sierra Nevada, California and Nevada*. US Geological Survey Water Supply Paper, 1535-I.)

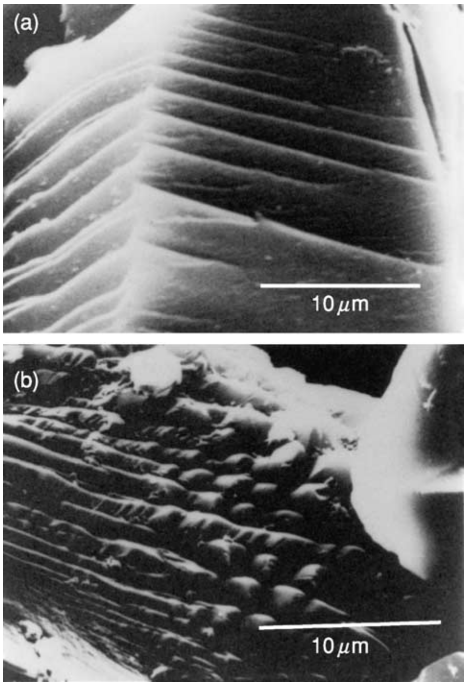


Figure 6 Scanning electron micrographs of an orthoclase particle treated with (a) water and (b) 0.01 mol l^{-1} oxalic acid. The mineral was treated for 20 days and the renewal treatments were then repeated twice for 5-day and 10-day periods. (Reproduced with permission from Song SK and Huang PM (1988) Dynamics of potassium release from potassium-bearing minerals as influenced by oxalic and citric acids. *Soil Science Society of America Journal* 52: 383–390.)

activity of each alkali ion to that of H ion. The diagrams show that mica and kaolinite are intermediate phases in the weathering of K-feldspar and that montmorillonite and kaolinite are intermediate phases in the weathering of albite. The ionic compositions of groundwater fall within the stability fields of kaolinite. This indicates that kaolinite is the stable mineral phase and mica and montmorillonite are unstable phases in groundwaters and eventually will be transformed to kaolinite. This is in agreement with the fact that groundwaters tend to transform rock minerals to kaolinite.

Feldspars are stores of the macronutrients K and Ca. Ligands of complexing organic acids may play a significant role in the release of these cations from the feldspar structure. Organic acids such as acetic, aspartic, salicylic, and citric acids have the ability to enhance the dissolution of feldspars. Their effect on dissolution increases in the order: acetic < aspartic < salicylic < citric. Citric acid is more effective than the other acids in releasing Ca, particularly from Ca-rich plagioclases, apparently because of the formation of Ca–organic complexes. In the case of the weathering of orthoclase by oxalic acid, the ridges on the cleavage surface (Figure 6) are evidently preferentially attacked by the acid and thus modified. The surfaces of orthoclase weathered by oxalic acid look much like those of naturally weathered K-feldspars. The sequence of K release from K-bearing minerals by oxalic and citric acids is biotite > microcline ≅ orthoclase > muscovite (Table 2). This differs from the sequence that occurs through cation exchange reaction

Table 2 K released from K-bearing minerals by various solutions at the end of a 10-day reaction period

	Organic acid[a]		
Mineral	Oxalic (g kg^{-1} structural K)	Citric (g kg^{-1} structural K)	NaTPB[b] (g kg^{-1} structural K)
Biotite	44	11	1000
Muscovite	1.1	0.9	83
Microcline	2.3	1.4	1.6
Orthoclase	2.2	1.2	1.3

[a] The concentration of organic acid solutions was 0.01 mol l^{-1}.

[b] 1 mol l^{-1} NaCl, 0.2 mol l^{-1} NaTPB, and 0.01 mol l^{-1} EDTA.

NaTPB, sodium tetraphenyl boron; EDTA, ethylenediaminetetraacetic acid.

Reproduced with permission from Song SK and Huang PM (1988) Dynamics of potassium release from potassium-bearing minerals as influenced by oxalic and citric acids. *Soil Science Society of America Journal* 52: 383–390.

using sodium tetraphenyl boron (NaTPB). The reaction of feldspars with complexing organic acids and its relation to nutrient release from soils merit attention.

Micas

Micas are 2:1 phyllosilicates with tightly held, nonhydrated, interlayer cations balancing a high layer charge. Primary attention is given to those micas with K$^+$ as the interlayer cation, because they are the abundant and important micas in most soils. Micas are abundant in many rocks, e.g., shales, slates, phyllites, schists, gneisses, and granites and in sediments derived from these and other rocks. The micas in soils are mainly inherited by soils from soil parent materials. Micas serve as precursors for expansible 2:1 phyllosilicates, i.e., vermiculites and smectites, to which micas may be transformed by replacement of the interlayer cations (usually K$^+$) by hydrated cations. Through K release, micas in soils are an important mineral source for plant growth.

The 2:1 layer of micas is composed of an octahedral sheet between two sheets of tetrahedra, as depicted in Figure 7. Micas can be classified as dioctahedral or trioctahedral, depending on the types and locations of cations in the octahedral sheets (Table 3). In trioctahedral micas such as biotite, all three octahedral positions are filled, whereas in dioctahedral micas such as muscovite, only two out of three octahedral cations are filled. Isomorphic substitution in the mica structure creates negative charge, which results in a strong coulombic attraction for charge-compensating interlayer cations such as K, which is not exchanged in the standard cation exchange-capacity determination. The negative layer charge of micas arises by some combination of three mechanisms: (1) substitution of R^{3+} (primarily Al or Fe) for Si^{4+} in tetrahedral positions, (2) substitution of R^{2+} for R^{3+} in octahedral positions, and (3) vacancies in octahedral positions. The resultant layer charge may originate entirely within the tetrahedral sheet, or may originate entirely within the octahedral sheet in some species, or may come partly from both sheets. The layer charge of micas is expressed on a formula unit basis (Table 3).

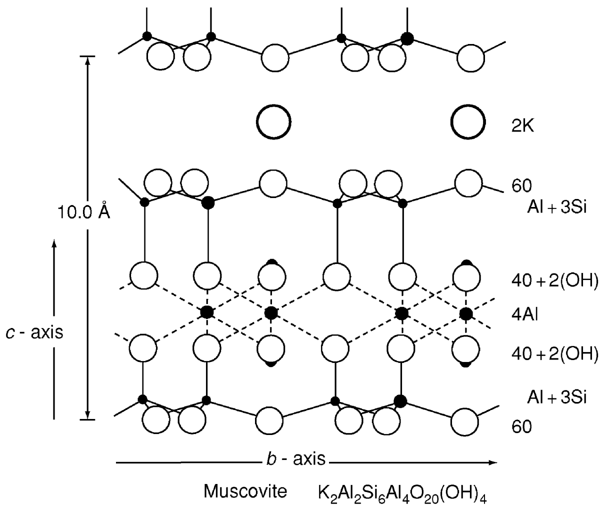


Figure 7 Schematic structure of mica. (Reproduced from Jackson ML (1964) Chemical composition of soils. In: Bear FE (ed.) *Chemistry of the Soil.* ACS Monograph Series 160. New York: Reinhold.)

Muscovite is the most abundant dioctahedral primary mica. In muscovite, excess negative layer charge results from the substitution of Al for Si in one out of each of four Si^{4+} in the tetrahedral sheet positions. Biotite is the most common trioctahedral mica. Approximately three-quarters of the cation locations in the tetrahedral sheets of biotite are filled with tetravalent Si and about one-quarter is filled with trivalent Al. The excess negative charge generated by substitution of Al for Si results in the negative charge that is localized in the basal O atoms of the tetrahedral sheets where Al substitution occurs. In soils and sedimentary rocks, mica in the clay fraction is usually identified as illite. Illitic minerals have been called 'hydrous mica,' 'clay mica,' and 'sericite.' Illite is commonly found to be dioctahedral, but its chemical composition and layer charge differ from that of muscovite (Table 3). Glauconite structure has many similarities to that of illite, with relatively little minor substitution of Al for Si in the tetrahedral sheet. Compared with illite, the octahedral sheet of glauconite contains considerably more Fe.

After feldspars and quartz, micas are the third most extensive group of minerals in granitic and sialic

Table 3 General chemical formulas and layer charge for some dioctahedral and trioctahedral micas

Mineral	Layer charge (mol per formula unit)	Cations: Tetrahedral sheet composition	Cations: Octahedral sheet composition	Cations: Interlayer cation	Anions
Dioctahedral					
Paragonite	1	Si_3Al	Al_2	Na	$O_{10}(OH)_2$
Muscovite	1	Si_3Al	Al_2	K	$O_{10}(OH)_2$
Phengite	1	$Si_{3.5}Al_{0.5}$	$Al_{1.5}[MgFe^{(II)}]_{0.5}$	K	$O_{10}(OH)_2$
Illite	0.75	$Si_{3.5}Al_{0.5}$	$Al_{1.75}[MgFe^{(II)}]_{0.25}$	$K_{0.75}$	$O_{10}(OH)_2$
Glauconite	0.8	$Si_{3.67}Al_{0.33}$	$[AlFe^{(III)}]_{1.33}[MgFe^{(II)}]_{0.67}$	K	$O_{10}(OH)_2$
Margarite	2	Si_2Al_2	Al_2	Ca	$O_{10}(OH)_2$
Trioctahedral					
Phlogopite	1	Si_3Al	Mg_3	K	$O_{10}(OH)_2$
Biotite	1	Si_3Al	$Mg_{0.6-1.8}Fe^{(II)}_{1.2-2.4}$	K	$O_{10}(OH)_2$
Annite	1	Si_3Al	$Fe^{(II)}_3$	K	$O_{10}(OH)_2$
Lepidolite	1	Si_3Al	$[Li,Al]_3$	K	$O_{10}(OH,F)_2$
Clintonite	2	$SiAl_3$	Mg_2Al	Ca	$O_{10}(OH)_2$

Adapted from Bailey SW (1984) Classification and structures of the micas. In: Bailey SW (ed.) *Reviews in Mineraology, Micas*, vol. 13. Washington, DC: Mineralogical Society of America, with permission.

(acid) rocks in general but less extensive in most mafic rocks. On land areas of the world, the sialic rocks are more extensive than mafic rocks. Muscovite and biotite are the most extensive micas in igneous and metamorphic rocks. Phlogopite occurs as a product of metamorphism of magnesian limestones or dolomitic limestones and also in serpentine rocks. Micas are generally more extensive in fine-grained sediments and sedimentary rocks (e.g., shales) than in coarse-textured sedimentary rocks (e.g., sandstones). Shales and slates are usually rich in the fine-grained, illitic-type micas. Illitic-type micas are also important minerals in limestones. Much illite in sedimentary rocks is apparently detrital. However, some illite in these rocks may also have been developed or affected by authigenic or diagenetic processes. The contents and species of micas in more recent, unconsolidated parent materials such as glacial tills, loess, and alluvium vary with their origin.

Micas represent a major source of the macronutrient K and thus play an important role in plant nutrition. If the soil solution is low in K^+ activity, K near the edges of micaceous minerals may be released into solution by exchange with H^+, Na^+, Mg^{2+}, or Ca^{2+}. When the readily exchangeable K is exhausted, fixed K from the interlayer position can also be released. Estimating the power of micaceous minerals to supply K to the soil solution is thus critical for predictions of plant nutrition and growth.

Micas exhibit a strong preference for Cs^+. This strong affinity of micas for large, weakly hydrated cations such as Cs^+ has an important environmental application in determining the fate of some radioactive waste. Radioactive isotopes of Cs have been added to soils by disposal of radioactive wastes and by deposition of fallout from both nuclear weapon testing and accidents at nuclear energy facilities. The mobility and bioavailability of radiocesium in mineral soils is limited by nearly instantaneous sorption of $^{137}Cs^+$ at the frayed edges of mica particles followed by slower fixation of $^{137}Cs^+$ ions diffused into interlayer sites within mica crystals. Therefore, micas play a role in restricting the contamination of $^{137}Cs^+$ to the terrestrial food chain and the impact on human health.

Structural Fe contents of micas may be substantial. For example, Fe accounts for approximately 23% by mass of ideal biotite. Structural Fe can participate in electron-transfer reactions in soils. Three examples of mica involvement in environmentally significant electron-transfer reactions are the reduction of NO_3^- to NH_4^+, the reduction of Cr(VI) to Cr(III), and the reductive degradation of chlorinated hydrocarbons in soils and groundwater.

Common Accessory Primary Minerals

Besides quartz, feldspars, and micas, other common primary minerals are olivines, pyroxenes, and amphiboles. These minerals are accessory primary minerals, present in the sand- and silt-sized fractions of soils.

Olivines occur as the principal constituents of some igneous rocks. In the upper lithosphere, however, their concentration is certainly smaller than that of pyroxenes and amphiboles. These three groups of minerals belong to the accessory minerals of soils and are present in heavy specific-gravity fractions.

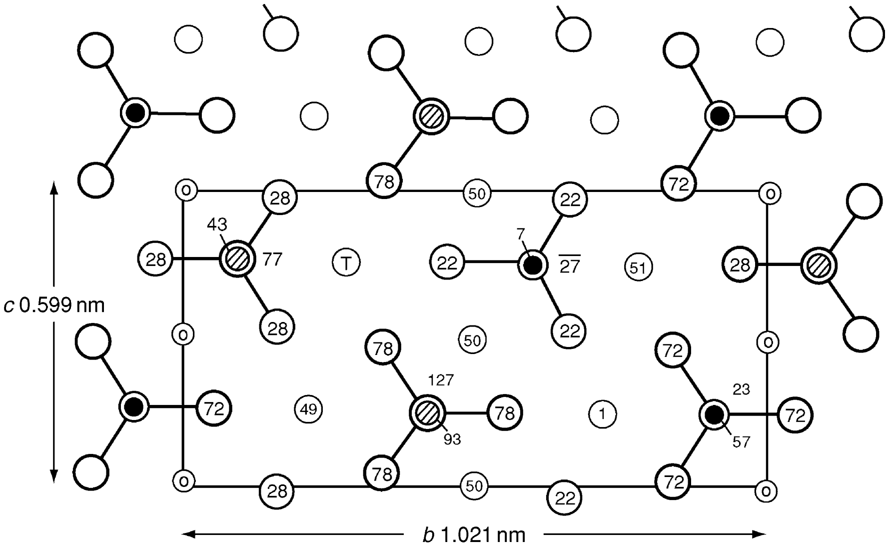


Figure 8 The structure of forsterite, Mg_2SiO_4, projected on the (100) plane. The largest circles are O, the smallest circles Mg, and the black or shaded circles Si; Si-O bonds mark out the SiO_4 tetrahedra. The kinds of positions of atoms are indicated by heights as shown by the numbers inside or beside the circles. (Reproduced from Bragg L and Claringbull GF (eds) (1965) *Crystal Structures of Minerals: The Crystalline State*, vol. IV. London, UK: G Bell.

Table 4 Certain members of olivine, pyroxene, and amphibole groups

Mineral	*Specific chemical formula*[a]
Olivines	
Forsterite	$(Mg_{0.96}Fe_{0.04})_2SiO_4$
Fayalite	Fe_2SiO_4
Tephroite	Mn_2SiO_4
Monticellite	$CaMgSiO_4$
Pyroxenes	
Enstatite	$MgSiO_3$
Hypersthene	$(Mg_{0.88}Fe_{0.12})Si_2O_6$
Orthoferrosilite	$FeSiO_3$
Clinoenstatite	$MgSiO_3$
Clinoferrosilite	$FeSiO_3$
Diopside	$CaMgSi_2O_6$
Pigeonite	$(Ca_{0.04}Mg_{0.45}Fe_{0.48})SiO_3$
Hedenbergite	$CaFeSi_2O_6$
Johannsenite	$(Ca_{0.45}Mg_{0.03}Fe_{0.05}Mn_{0.47})SiO_3$
Acmite-augite	$(Ca, Na)(Fe, Mn, Zn, Mg)(Si, Al)_2O_6$
Aegirine (acmite)	$NaFeSi_2O_6$
Jadeite	$NaAlSi_2O_6$
Spodumene	$LiAlSi_2O_6$
Amphilboles	
Anthophyllite	$(Mg, Fe)_7Si_8O_{22}(OH)_2$
Cummingtonite	$(Fe_{0.6}Mg_{0.4})_7Si_8O_{22}(OH)_2$
Tremolite[b]	$Ca_2Mg_5Si_8O_{22}(OH)_2$
Arfvedsonite	$(Ca, Na, K)_{2.64}Fe_{1.42}(Fe, Mn, Mg, Ti)_{3.54}(Si, Al)_8O_{22}(OH)_{2.15}$
Glaucophane	$Na_2(Mg, Fe, Al)_5Si_8O_{22}(OH)_2$
Riebeckite	$(Na, Ca, K)_{2.2}(Fe^{2+}, Mg, Mn, Zn, Cu, Li)_{3.0}(Fe^{3+}, Al, Ti)_{2.0}(Si, Al)_{8.0}O_{22.1}(OH, F)_{1.9}$
Hornblende	$(Ca, Na)_{2.26}(Mg, Fe, Al)_{5.15}(Si, Al)_8O_{22}(OH)_2$

[a] The specific chemical formulas are compiled from the *Data Book* (after Joint Committee on Powder Diffraction Standards (1974) *Search Manual. Selected Powder Diffraction Data for Minerals, Data Book*, 1st edn. Swarthmore, PA: International Center for Diffraction Data).

[b] The division between tremolite and actinolite is arbitrary: tremolite is the low-iron end of the series and actinolite comprises the high-iron members.

Reproduced with permission from Huang PM (1989) Feldspars, olivines, pyroxenes, and amphiboles. In: Dixon JB and Weed SB (eds) *Minerals in Soil Environments*. Madison, WI: Soil Science Society of America.

Weathering of these minerals is one of the major processes in the geochemical cycles of Mg and Fe. The variety of isomorphous substitution possible in these minerals and their relative ease of weathering make them excellent source minerals for Ca, Mg, and trace elements in soils. Furthermore, the catalytic power of Mn-bearing nesosilicates and inosilicates in the abiotic formation of humic substances in soils merits attention.

Olivines

Members of forsterite-fayalite series, $(Mg, Fe)_2SiO_4$, comprise the most abundant naturally occurring olivines. The structure of forsterite is shown in Figure 8. The O atoms lie in approximately hexagonally close-packed sheets parallel to (100). The SiO_4 tetrahedra point alternately in opposite directions along the *a*- and *b*-axis. Every Mg atom is octahedrally coordinated by six O atoms. In the structures described above, Fe^{2+} may replace Mg^{2+}, because the two have identical charge and similar ionic radii. The specific chemical formulas of selected minerals in the olivine group are given in Table 4.

There is complete solid solution between forsterite and fayalite in nature. The Mg-rich olivines are predominant in basic rocks and the Fe-rich varieties in intermediate and acidic igneous rocks. Tephroite occurs mainly in metasomatic rocks and in Mn deposits.

The weatherability of olivines in soils is very high. Due to their relatively easy weatherability, olivines contribute to nutrient levels of Mg and Fe. The extent of weathering of these minerals is an indication of the intensity of chemical weathering of soils in cool humid regions. Where leaching is only moderate, olivines may alter to serpentine or trioctahedral smectite (presumably of the saponite type), nontronite, and various ferric hydrates and gels. In an environment where leaching is intense, the degradation products are poorly crystallized smectite, kaolinite,

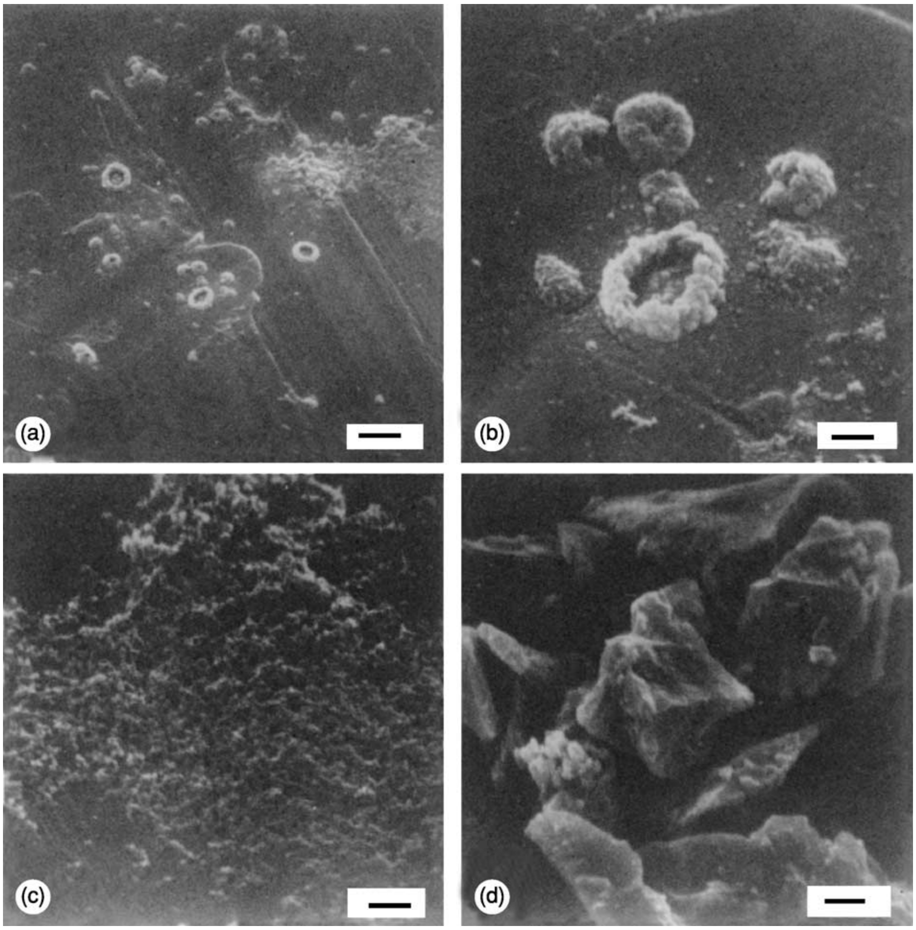


Figure 9 Scanning electron microscope micrographs of hydroquinone polymers in the supernatant and mineral particles settled in the tephroite system at the ratio of mineral to hydroquinone solution of 0.01 at the initial pH of 6.0 at the end of 7 d. (a–c) Hydroquinone polymers; (d) tephroite particles after reaction with hydroquinone. (a) Bar scale 10 μm; (b–d) 2 μm. (Reproduced with permission from Shindo H and Huang PM (1985) Catalytic polymerization of hydroquinone by primary minerals. *Soil Science* 139: 505–511.)

halloysite, and Fe oxides such as goethite, hematite, or the noncrystalline precursors of these minerals.

Olivines have the ability to accelerate the abiotic polymerization of hydroquinone to humic substances. The catalytic power of tephroite, which is an Mn-bearing olivine, in the polymerization of polyphenols such as hydroquinone to humic acids with a high degree of humification is especially noteworthy. The scanning electron micrographs of the hydroquinone-derived humic polymers formed in the presence of tephroite reveal various morphological configurations (Figure 9a–c). The surface features of these polymers are similar to those of soil humic and fulvic acids. Furthermore, the infrared (IR) spectra of these humic polymers with molecular weights of approximately 3500 and greater are similar to those of humic substances. The association of hydroquinone-derived polymers with the surfaces of tephroite is not evident (Figure 9d), indicating that most humic polymers formed in the tephroite system are present as discrete particles.

Pyroxenes and Amphiboles

Pyroxenes and amphiboles are closely related structurally. Pyroxenes consist of single chains of linked SiO_4 tetrahedra, each of which shares two O atoms with its neighbor; in amphiboles, double chains of the tetrahedra are linked together by O atoms and the tetrahedra share alternately two and three O atoms (Figure 10). The chains are indefinite in length. When viewed at right angles to the chain and parallel to the base of the tetrahedra, pyroxenes and amphibole chains look alike, as depicted in Figure 10c. The pyroxene and amphibole chains are parallel to the *c*-axis of the crystal and their planes of basal O atoms are perpendicular to (010).

The chains of pyroxenes and the double chains of amphiboles are linked together by various cations. The general chemical formula for pyroxenes is $R_2[Si_2O_6]$ and that for amphiboles is $R_{14}[(OH)_4 Si_{16}O_{44}]$. In these formulas, R is Mg, Fe^{2+}, or Ca and, in many cases, A1, Fe^{3+}, Ti^{3+}, Mn^{3+}, Na, K, or Li. The bond between the O atoms and the cations linking the chains is relatively weaker than that between the O and the Si. Therefore, cleavage takes place diagonally through the crystal and does not rupture the Si-O chains. The principal members of pyroxene and amphibole groups are listed in Table 4.

Three series of pyroxenes – the enstatite-hypersthene, diopside-jadeite, and augite series – occur in

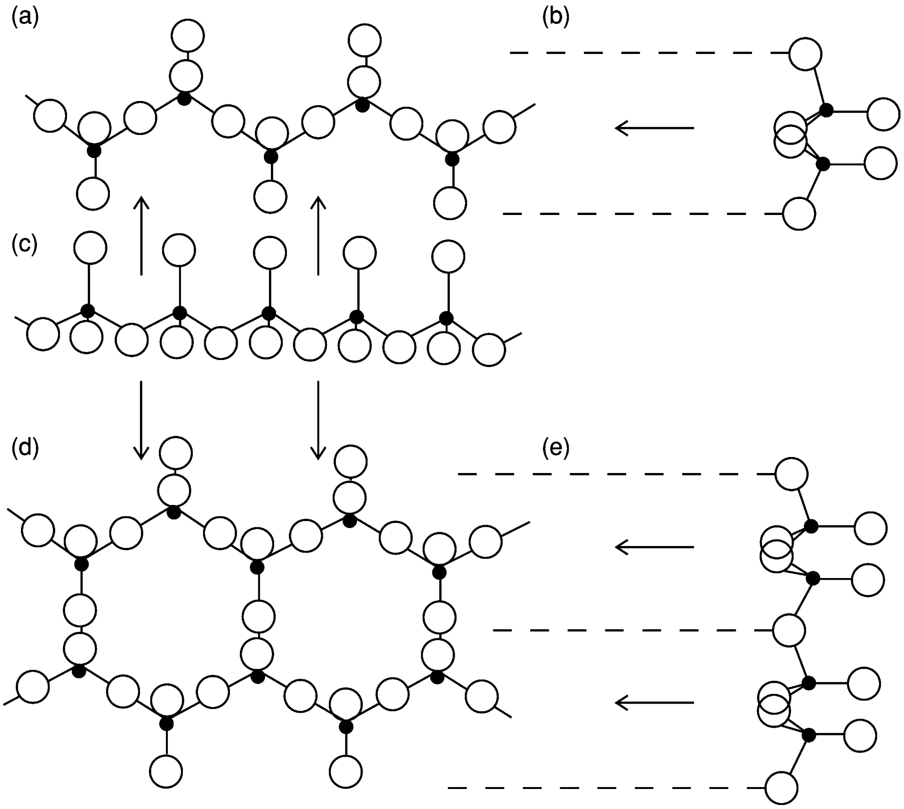


Figure 10 Si-O chains in structures of pyroxenes and amphiboles: (a–c) the pyroxene chain as seen in plane, in elevation, and end-on view, respectively; (c–e) the amphibole chain from the same three aspects. (Reproduced from Bragg L and Claringbul GF (eds) (1965) *Crystal Structures of Minerals: The Crystalline State*, vol. IV. London, UK: G Bells and Sons, Ltd.)

igneous rocks. In nature, pyroxenes are often mixtures. Among the three series of pyroxenes, the augite series is geochemically more important than the others. Its members are among the most important constituents of igneous rocks. Many amphiboles are absent in primary igneous rocks or may occur as minerals of secondary origin. The main primary amphiboles of igneous rocks are the Mg-poor riebeckites and the Si-poor basaltic and common hornblende. Hornblendes are widely distributed in igneous rocks from syenite and granite to gabbro, and in metamorphic rocks such as gneiss, hornblende schist, and amphibolites. Pyroxenes and amphiboles are largely confined to the sand and silt fractions of soils. Some of them may occur in the clay fraction of soils that are developed from glacial rock flour and have not been subjected to intensive weathering. The variety of isomorphous substitution in these minerals and their easy weatherability make them excellent sources of trace elements and Ca and Mg in soils.

Pyroxenes tend to weather to chlorite or smectite or both, through partial dissolution of Mg, Ca, and Fe^{2+}. Calcite may develop in the environment where the rate of dissolution of Ca is higher than the rate of complete breakdown of the pyroxene. As the chemical weathering proceeds further, all the Ca and Mg, and part of the Si are lost, resulting in progressive enrichment of the residue in kaolinite, ferric oxides and oxyhydroxides, and anatase. Amphiboles appear to have a weathering sequence similar to pyroxenes.

Pyroxenes (e.g., augite) and amphiboles (e.g., hornblende) have the ability to catalyze abiotic polymerization of polyphenols such as hydroquinone to humic substances. Therefore, these minerals along with olivines merit attention in the abiotic formation of humic substances in soil environments. These accessory primary minerals thus play a role in catalyzing humification and protecting soil C against mineralization and loss to the atmosphere as CO_2.

See also: **Factors of Soil Formation:** Parent Material; **Mineral–Organic–Microbial Interactions**

Further Reading

Bailey SW (1984) Classification and structures of the micas. In: Bailey SW (ed.) *Micas. Reviews in Mineralogy,* vol. 13, pp. 1–2. Washington, DC: Mineralogical Society of America.

Berry LG and Mason B (1959) *Mineralogy.* San Francisco, CA: WH Freeman.

Bloom PR and Nater EA (1991) Kinetics of dissolution of oxides and primary silicate minerals. In: Sparks DL and Suarez DL (eds) *Rates of Soil Chemical Processes*, pp. 151–189. SSSA Special Publication 27. Madison, WI: Soil Science Society of America.

Bragg L and Claringbull GF (1965) *Crystal Structure of Minerals: The Crystalline State,* vol. IV. London, UK: G Bell.

Churchman GJ (2000) The alteration and formation of soil minerals by weathering. In: Sumner ME (ed.) *Handbook of Soil Science*, pp. F3–F76. Boca Raton, FL: CRC Press.

Drees LR, Wilding LP, Smeck NE, and Senkayi AL (1989) Silica in soils; quartz and disordered silica polymorphs. In: Dixon JB and Weed SB (eds) *Minerals in Soil Environments*, pp. 914–974. SSSA Book Series 1. Madison, WI: Soil Science Society of America.

Fanning DS, Keramidas VZ, and El-Desoky MA (1989) Micas. In: Dixon JB and Weed SB (eds) *Minerals in Soil Environments*, pp. 551–634. SSSA Book Series 1. Madison, WI: Soil Science Society of America.

Feth JH, Robertson CE, and Polzer WL (1964) *Sources of Mineral Constituents in Water from Granitic Rocks, Sierra Nevada, California and Nevada*. US Geological Survey Water Supply Paper, 1535-I. Washington, DC: US Department of the Interior, US Geological Survey.

Huang PM (1989) Feldspars, olivines, pyroxenes, and amphiboles. In: Dixon JB and Weed SB (eds) *Minerals in Soil Environments*, pp. 975–1050. SSSA Book Series 1. Madison. WI: Soil Science Society of America.

Jackson ML (1964) Chemical composition of soils. In: Bear FE (ed.) *Chemistry of the Soil*, pp. 71–141. ACS Monograph Series No. 160. New York: Reinhold.

Laves FJ (1960) A1/Si-Verteilungen, Phasen-Transformationen und Namen der Alkali-feldspäte. *Zeitschrift fur Kristallographie, Kristallgeometrie, Kristallphysik, Kristallchemie* 113: 265–296.

Shindo H and Huang PM (1985) Catalytic polymerization of hydroquinone by primary minerals. *Soil Science* 139: 505–511.

Somasiri S and Huang PM (1973) The nature of K-feldspars of selected soils in the Canadian prairies. *Soil Science Society of America Proceedings* 37: 461–464.

Song SK and Huang PM (1988) Dynamics of potassium release from potassium-bearing minerals as influenced by oxalic and citric acids. *Soil Science Society of America Journal* 52: 383–390.

Taylor WH (1965) Framework silicates: the feldspars. In: Bragg L and Claringbull GF (eds) *Crystal Structures of Minerals: The Crystalline State,* vol. IV, pp. 293–333. London, UK: G Bell.

Minerals, Secondary *See* **Clay Minerals**

Minimum Tillage *See* **Conservation Tillage**

Miscible Displacement *See* **Solute Transport**

MORPHOLOGY

P R Owens, USDA Agricultural Research Service, Mississippi State, MS, USA
E M Rutledge, University of Arkansas, Fayetteville, AR, USA

Introduction

Soil morphology is defined as the branch of soil science that deals with the description, using standard terminology, of *in situ* spatial organization and physical properties of soil regardless of potential use. Precise descriptions, using conventional terms, are necessary to all of the areas of science included in the field of soil science. The basic descriptors used today have been developed over the past 50 years and are continually added to. Prior to utilization of standard terminology, soils were described as clayey, sandy, stony, sedimentary, saline, marshy, dry or moist, heavy or light, soft or compact, fatty, friable, or lean. Soil horizons were described by terms such as 'gray watery sand' or 'rusty brown clay.' While these terms illustrate a fundamental understanding of soil properties, they do not impart any specific knowledge about the soils in question and cannot be compared with other soils described by other scientists. An objective, complete description of the soil is essential, because it serves as a basis for soil identification, classification, correlation, mapping, and interpretation.

The basic tools used by field soil morphologists are very simple and usually include a knife, tape measure, water bottle, Munsell color book, ×10 hand lens, 10% HCl, and a clinometer or Abney level. A description of a soil profile, a vertical cross-section of the soil, represents a three-dimensional pedon. A pedon is the smallest volume that is recognized as a soil. Field morphology of a soil profile is best determined from a recently opened pit; however, a fresh roadside cut may be used. Field descriptions begin by organizing and subdividing a vertical exposure (soil profile) into reasonably distinct layers or horizons. At a minimum, a soil description includes a site characterization and the horizon depth intervals, horizon boundary characteristics, color, texture, structure, consistence, roots and pores, pH, and effervescence; and special descriptions of special features for each horizon.

Site Description

The soil pedons that are described at a site will have been affected by five soil-forming factors: parent material, biota, climate, topography, and time. A brief description of the parent material is desirable, because of inherited soil properties. It is necessary to record the observable features in the landscape that impart information about a soil. The geomorphic information needs to be described as precisely as possible, including the physiographic location, geomorphic description, and surface morphometry. The physiographic information includes: (1) physiographic division; (2) physiographic province; (3) physiographic section; (4) state physiographic area; and (5) local physiographic or geographic name. The geomorphic information includes: (1) landscape, (2) landform, (3) microfeatures, and (4) anthropogenic features. The surface morphometry includes: (1) elevation; (2) slope aspect; (3) slope gradient; (4) slope complexity; (5) slope shape; (6) geomorphic components; and (7) microrelief. The vegetation type is described, listing the common names of the plants and their scientific names when possible. The latitude and longitude, in degrees, seconds, and minutes, are most commonly used to record the geographical location of the soil as precisely as possible.

Horizons

Soil horizons may be the most distinctive visual features of a soil profile. According to the *Soil Survey Manual*, a soil horizon is a layer approximately parallel to the surface of the soil, distinguishable from adjacent layers by a distinctive set of properties produced by the soil-forming processes. The term 'layer,' rather than 'horizon,' is used if all the properties are believed to be inherited from the parent material or no judgment is made as to whether the layer is genetic.

The depth to and thickness of the horizon should be recorded as well as the horizon designation. Horizons are labeled according to diagnostic features and interpretations. A master horizon label is given which represents characteristic properties. As an example, the system developed by the US Department of Agriculture (USDA) uses O, A, E, B, C, and R as master horizon labels. These labels correspond to:

1. O horizons and layers: Layers dominated by organic material. Some are saturated with water for long periods or were once saturated but are now artificially drained; others have never been saturated.
2. A horizons: Mineral horizons that formed at the surface or below an O horizon which exhibit obliteration of all or much of the original rock structure and that show one or more of the following: (1) an accumulation of humified organic matter intimately mixed with the mineral fraction and not dominated by properties characteristic of E or B horizons; or (2) properties resulting from cultivation, pasturing, or similar kinds of disturbance.
3. E horizons: Mineral horizons in which the main feature is loss of silicate clay, iron, aluminum, or some combination of these, leaving a concentration of sand and silt particles.
4. B horizons: Horizons that formed below an A, E, or O horizon, contain soil structure, and show one or more of the following: (1) illuvial concentration of silicate clay, iron, aluminum, humus, carbonates, gypsum, or silica, alone or in combination; (2) evidence of the removal of carbonates; (3) residual concentration of sesquioxides; (4) coatings of sesquioxides that make the horizon conspicuously lower in value, higher in chroma, or redder in hue than overlying and underlying horizons without apparent illuviation of iron; (5) alteration that forms granular, blocky, or prismatic structure if volume changes accompany changes in moisture content; or (6) brittleness.
5. C horizons or layers: Horizons or layers, excluding hard bedrock, that are little affected by pedogenic processes and lack properties of O, A, E, or B horizons. The material of C layers may be either like or unlike that from which the solum presumably formed.
6. R layers: Hard bedrock.

Soil horizons can be dominated by properties of one master horizon but have subordinate properties of another; in which case, the master horizons may be combined to denote this interpretation (i.e., AB, EB, CB, BC, etc.). Where soil material from two horizons is within the same horizon, a virgule (/) is used to indicate two materials within that same horizon that can be described separately with regard to color, texture, and structure (i.e., Bw/A, Bt/A, C/Bt, etc.).

Designations of soil horizons are supplemented with suffixes or subscripts to convey more information about a horizon. Description systems vary slightly; however, as an example the suffixes used by the USDA are listed in Table 1.

Soil horizons and layers have been developed based on central concepts observed from field investigations. The properties between soils vary greatly and are a continuum of properties. There are no definitions that are rigorously followed to label each horizon. The labeling is an interpretation based on observations of the scientist; however, the classification of the soil, which is based on the description, does follow criteria definitions.

Table 1 Suffix symbols used by the USDA for master horizons

Suffix	*Horizon*
a	Highly decomposed organic material
b	Buried genetic horizon
c	Concretions or nodules
d	Physical root restriction
e	Organic material of intermediate decomposition
f	Frozen soil or water
ff	Dry permafrost
g	Strong gleying
h	Illuvial accumulation of organic matter
i	Slightly decomposed organic material
j	Accumulation of Jarosite
jj	Evidence of cryoturbation
k	Accumulation of carbonates
m	Cementation or induration
n	Accumulation of sodium
o	Residual accumulation of sesquioxides
p	Tillage or other disturbance
q	Accumulation of silica
r	Weathered or soft bedrock
s	Illuvial accumulation of sesquioxides and organic matter
ss	Presence of slickensides
t	Accumulation of silicate clay
v	Plinthite
w	Development of color or structure
x	Fragipan character
y	Accumulation of gypsum
z	Accumulation of salts more soluble than gypsum

Reproduced from Soil Survey Staff (1998) *Keys to Soil Taxonomy*, 8th edn. USDA, Natural Resources Conservation Service. Washington, DC: US Government Printing Office.

Soil Texture

Soil texture is one of the most common features used by scientists and laymen to describe soils. Texture refers to the relative abundance, by weight, of the size fractions – sand, silt, and clay. In all textural classification systems, the sum of all particles (by weight) less than or equal to 2 mm is equal to 100% (Figure 1), and is referred to as the 'fine-earth fraction.' There are several textural classification systems used, with varying ranges of particles sizes (Figure 2).

Textural descriptions are commonly used, because it is cumbersome to give the percentage of sand, percentage of silt, and percentage of clay within a description. Commonly, field estimates of sand, silt, and clay are used, and the soils are placed into textural classes. These textures are verified by laboratory analysis. The mineralogy of the clay may affect estimates of clay in field textures. The kaolinitc clays expand less and tend to be less expressed than smectitic clays which may result in an underestimation of the percentage of clay.

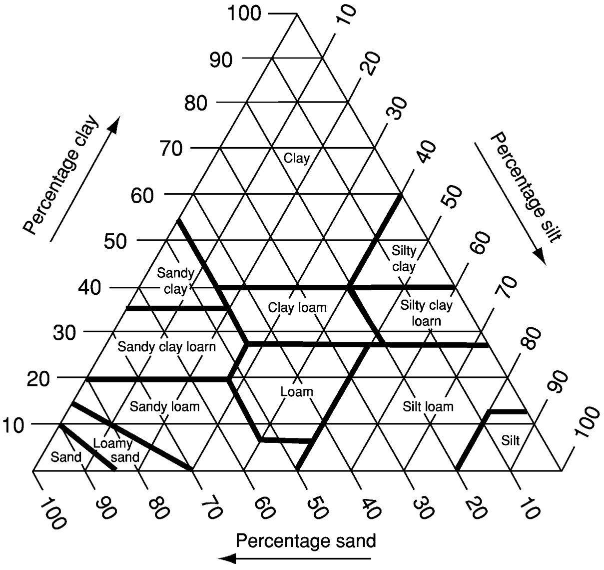


Figure 1 USDA soil textural triangle.

Soil texture relates to many of the ways a soil performs. If a soil is coarse (sandy), water tends to move through it quite well, but it may not retain enough water for plant growth. If a soil is clayey, water will probably move slowly but it should retain water for plant growth. The shrink–swell potential of clayey soils varies quite widely depending on the type of clay minerals present. If the soil is high in smectitic clay minerals, it will shrink and swell as the moisture content of the soil changes. If the shrinking and swelling are not considered in the design of infrastructure, it will cause problems for roads and foundations. However, if the clay mineral is primarily kaolinite and other low-activity minerals, construction on such soils should not be a problem. Also, if the particles larger than 2 mm are of sufficient quantity to be important, the textural classes are prefixed with modifiers. These separates are classed as gravel, cobbles, or stones (Figure 2). (*See* **Texture**.)

Soil Color

Soil colors are used to infer pedogenic processes in soils. The main pigmenting (coloring) agents in soils are organic matter, iron, and, to a lesser extent, manganese. When these agents are not covering the mineral grains, the natural color of the grains is visible. Most mineral grains are naturally gray. The contrast of color is shown in Figures 3 and 4, where two soils

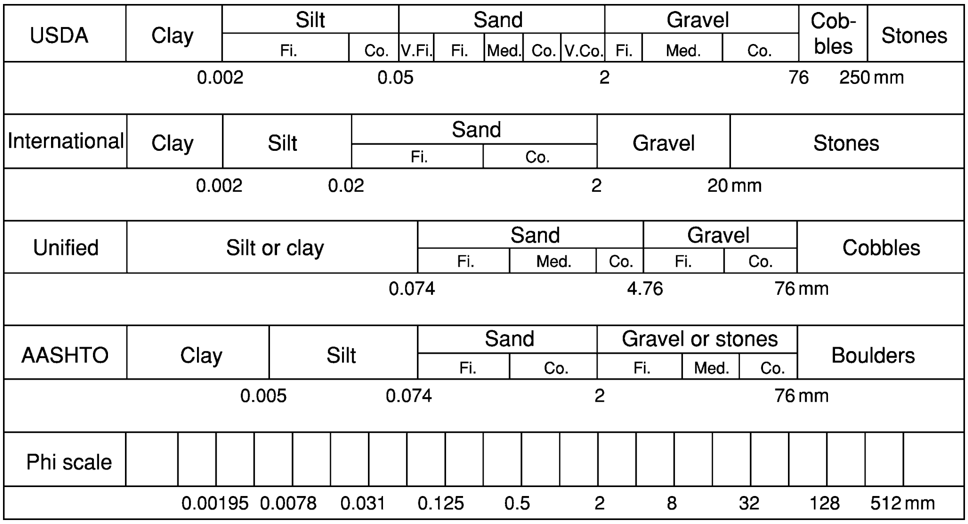


Figure 2 Particle-size classes using five particle-size class systems. Fi., fine; Co., coarse; V.Fi., very fine; V.Co., very coarse; Med., medium.

Figure 3 Highly developed soil in a well-drained convex position.

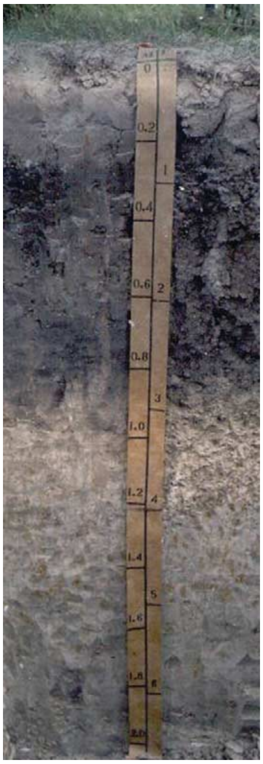


Figure 4 Highly developed soil in a poorly-drained concave position.

have similar texture and structure yet differ in color scheme. When a soil horizon has more than one color, the dominant color by volume is the matrix color.

Soil colors are most conveniently measured by comparison with a soil color chart. The one generally used is a modification of the Munsell color chart that includes only the portion needed for soil colors, about one-fifth of the entire range of color. The arrangement is by hue, value, and chroma – three simple variables combine to give all colors. Hue is the dominant spectral (rainbow) color; it is related to the dominant wavelength of the light. Value refers to the relative lightness of the color and is a function (approximately the square root) of the total amount of light. Chroma is the relative purity or strength of the spectral color and increases with decreasing grayness.

In the soil color chart, all colors on a given card are of a constant hue, designated by a symbol in the upper right-hand corner of the card. Vertically, the colors become successively lighter by visually equal steps; their value increases. Horizontally, they increase in chroma to the right and become grayer to the left. The value and chroma of each color in the chart are printed immediately beneath the color. The first number is the value and the second is the chroma. As arranged in the chart, the colors form three scales: (1) radial, or from one card to the next, in hue; (2) vertical in value; and (3) horizontal in chroma.

The matrix color is the first color described in a soil horizon or layer, since it is by definition the most dominant color. The moisture status of the soil color must be taken into account when describing soil color, i.e., moist or dry. To determine soil color, a fresh ped (a unit of soil structure) interior is used to eliminate the possibility of describing the color of a ped surface coating.

Redoximorphic Features and Mottles

The lack of uniformity of soil color may give insight regarding the pedogeneis of the soil. In the USDA description terminology, the generic term 'mottle' was changed to 'redoximorphic feature' when describing features in the soil resulting from reduction–oxidation reactions. The term 'redoximorphic feature' is a more descriptive term used when the pedogenic processes of reduction and oxidation are interpreted. 'Mottles' is a term used for spots in the soil and infers no pedogenic process.

Redoximorphic Features

Soil colors are greatly affected by the oxygen supply in the soil. As long as there is sufficient oxygen for microbial activity, the soils are usually brown or red and frequently have a single dominant color.

However, when soils become wet (saturated with water), the oxygen supply cannot flow as freely and the supply becomes insufficient for many microbes. Insufficient oxygen in the soil results in reduction – a process by which iron and manganese gain electrons.

Figure 5 Iron depletions illustrated in the subsoil.

Reducing conditions in soils are precursors to redoximorphic features, which are multicolored. Reduction of iron and manganese greatly increases their solubility, thus allowing them to dissolve into solution and move with the water, i.e., become mobile. Features caused by reduction or reduction and subsequent reoxidation are also called 'redoximorphic features.'

When the iron in the soil goes into solution, it uncoats the mineral grains, resulting in zones called 'depletions,' where the color is that of the mineral grains – usually gray. When air reenters the soil, the iron and manganese are oxidized and precipitate into brownish to reddish to blackish forms called 'concentrations.' Thus depletions are caused by removal of iron coatings, and concentrations are caused by oxidation and precipitation of previously reduced forms of iron and manganese. Both depletions and concentrations require reduction, but only concentrations require subsequent reoxidation.

Concentrations and depletions are described by kind, size, abundance, and color. Depletions are usually lighter colored than the matrix (Figure 5), and concentrations are usually browner, redder, or blacker than the matrix (Figure 6). There are several kinds of concentrations, which are described as finely disseminated materials, masses (noncemented), nodules, or concretions (cemented). Masses are noncemented concentrations of iron and manganese. Concretions and nodules are both cemented accumulations of iron and manganese, but differ in their internal structure: concretions have concentric rings, whereas nodules do not (Figure 7). There are two types of depletions: (1) iron depletions, which are primarily a lighter color

Figure 6 Iron concentrations.

Figure 7 A manganese nodule in a loess soil.

than the matrix; and (2) clay depletions, which have the same color change but also have lost clay and thus are grainier.

Mottles

Mottles (areas of color that differ from the matrix color) are commonly lithochromic or lithomorphic colors. Commonly these features are inherited from the geologic parent material rather than pedogenesis. Also included in mottles are relic redoximorphic features, which are features that formed in a pluvial environment or in landscape that has been naturally or anthropogenically altered and permanently affects the drainage of the soil. Description of mottles may include abundance, size, contrast (to matrix), color, and name of object boundary.

Soil Structure

An undisturbed soil forms several levels of structure that can be divided into primary, secondary (compound), and tertiary structure. The primary structural units are the basic structural units used in morphological descriptions. Secondary and tertiary structure are defined as: (1) the size, shape, and arrangement of the primary peds, their interpedal voids, and associated interpedal pedological features in a soil material; (2) the size, shape, and arrangement of the secondary peds of a sol material (compound peds resulting from the packing of primary peds), their interpedal voids and associated interpedal pedological features. Soil structure refers to units composed of primary particles. The cohesion within these units is greater than the adhesion among units and is a product of processes that aggregate, cement, compact, or unconsolidate soil particles. Under stress, the soil mass tends to rupture along predetermined planes or zones. The peds form repetitive, quasi-uniform shapes.

Clods and fragments stand in contrast to peds, for which soil-forming processes exert weak or no control on the boundaries. A clod is caused by disturbance, such as plowing or digging, which molds the soil to a transient mass that slakes with repeated wetting and drying. Fragments include: (1) units of undisturbed soil with bounding planes of weakness that are formed on drying without application of external force and which do not appear to have predetermined bounding planes; (2) units of soil disturbed by mechanical means but without significant rearrangement to a denser configuration; and (3) pieces of soil bounded by planes of weakness caused by pressure exerted during examination, with size and shape highly dependent on the manner of manipulation.

Some soils lack structure and are referred to as structureless. When structureless soils are ruptured, soil fragments, single grains, or both may result. Structureless soil may be termed either ‘massive’ or ‘single grain.’ Massive soils tend to form fragments when disturbed, and single grain soils are usually sandy and are loose in 50% or more of the mass.

Shape

Several basic shapes of structural units are recognized in soils. Supplemental statements about variations in shape of individual peds are needed in detailed descriptions of some soils. Table 2 and Figure 8 describe the possible shapes of soil structure.

Table 2 Types of soil structure

Structure	*Description*
Natural soil structural units (pedogenic structure)	
Granular	Small peds, with rounded or very irregular faces
Angular blocky	Peds with faces that intersect at sharp angles
Subangular blocky	Peds with sub-rounded shapes and lacking sharp angles
Platy	Flat plate-shaped units
Wedge	Elliptical, interlocking lenses that terminate in acute angles, bounded by slickensides; not limited to vertic materials
Prismatic	Vertically elongated peds with flat tops
Columnar	Vertically elongated peds with rounded tops
Structureless	
Single grain	No structural units; entirely noncoherent; e.g., loose sand
Massive	No structural units; material is a coherent mass

Reproduced from Soil Staff Survey (1993) *Soil Survey Manual*. USDA Soil Conservation Service, Agricultural Handbook No. 18. Washington, DC: US Government Printing Office.

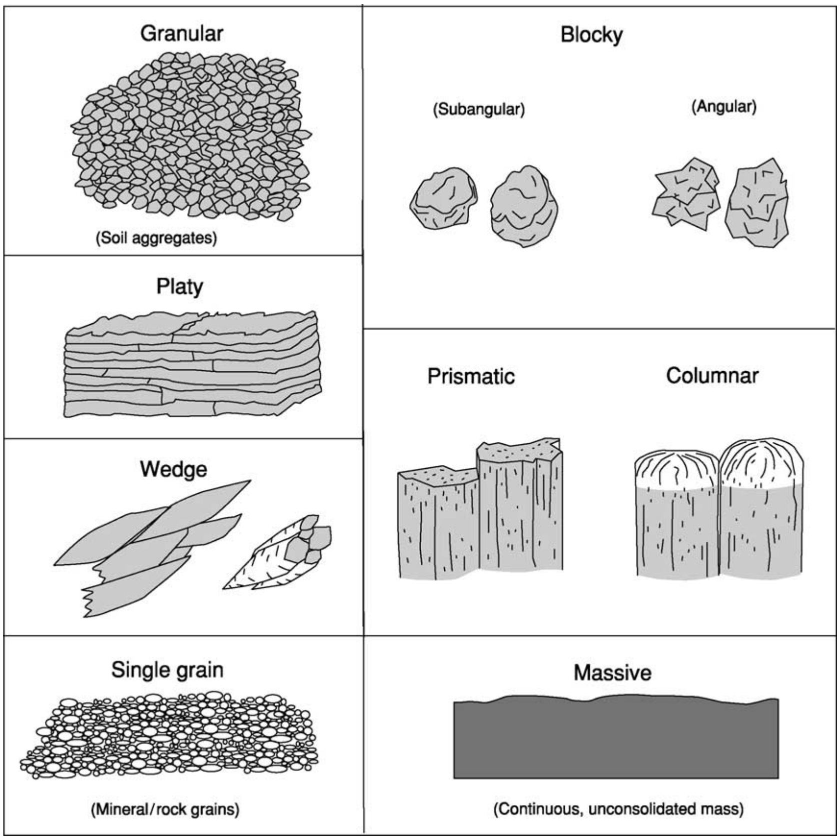


Figure 8 Shapes of soil structure. (Reproduced from Schoenberger PJ, Wysocki DA, Benham EC, and Broderson WD (eds) (2002) *Field Book for Describing and Sampling Soils*, version 2.0. Natural Resources Conservation Service. Lincoln, NE: National Soil Survey Center.)

Size

There are five classes that are used to describe the size of structural units: very fine, fine, medium, coarse, and very coarse. The size limits vary according to the shape of the unit. The size limits refer to the smallest dimension of plates, prisms, and columns.

Grade

The distinctness of units is described by the grade. Criteria are based on the ease of separation into discrete units and the proportion of units that hold together when the soil is handled.

Weak Units that are barely observable in situ are termed 'weak.' When gently disturbed, the soil material parts into a mixture of whole and broken units and much material that exhibits no planes of weakness. Faces that indicate persistence through wet–dry–wet cycles are evident.

Moderate Units graded as moderate are well formed and evident in undisturbed soil. When disturbed, the soil material parts into a mixture of mostly whole units, some broken units, and material that is not units. Peds part from adjoining peds to reveal nearly entire faces that have properties distinct from those of fractured surfaces.

Strong Units that are distinct in undisturbed soil are graded 'strong.' They separate cleanly when the soil is disturbed. When removed, the soil material separates mainly into whole units.

The distinctness of individual structural units and the relationship of cohesion within units to adhesion between units determine grade of structure. The degree of disturbance required to determine structure grade depends largely on moisture content and percentage and mineralogy of clay. Only slight disturbance may be necessary to separate the units of moist sandy loam having strong granular structure, while considerable disturbance may be required to separate units of a moist clay loam having strong blocky structure.

The designation of structure grade, size, and shape can be modified with other appropriate terms when necessary to describe other characteristics. Surface characteristics of units are described separately. Special structural units, such as wedge-shaped units of high-potential shrink–swell soils, are described in appropriate terms.

Compound Structure

'Secondary structure' is another term for 'compound structure,' where smaller structural units may be held together to form larger units. Grade, size, and shape are given for both, and the relationship of one set to the other is indicated. 'Strong medium blocks within moderate coarse prisms' and 'moderate coarse prismatic structure parting to strong medium blocky' are examples of compound structure.

Soil Consistence

Consistence refers to the cohesion and adhesion within a soil or its resistance to deformation or rupture. The consistence of soils changes greatly as the soil moisture changes. There are many expressions of consistence, such as rupture resistance, surface crust, manner of failure, stickiness, plasticity, penetration resistance, and others.

The most commonly used parameter of consistence is field evaluation (i.e., rupture resistance), which is conducted by placing a specimen 25–30 mm on each side between the thumb and first finger and applying pressure. Again, the moisture content of the soil sample is very important.

Soils in humid and subhumid climates may form brittle subsurface horizons. The degree and percentage of brittleness of the soil are described. Soils may also be hardened due to cementation. In this case, the materials accumulated causing the cementation, such as carbonates, gypsum, humus, iron, and/or silica are listed. In addition, the percentage, cementation of soil and the strength of the cementation are noted.

Roots and Pores

The roots and pores that can be observed in the field are among the larger pores in the soil. These pores are mostly noncapillary pores, i.e., pores that cannot hold water against the force of gravity. These pores are of major importance in moving water and air within the soil. After a major wetting of the soil (rainfall, snowmelt, irrigation, etc.), these pores must transmit the water throughout the soil to the smaller capillary pores. The noncapillary pores should redistribute the water in a few days, because they must drain in order also to distribute oxygen (air) into the soil. If soil oxygen is depleted for more than a few days, many plants will die or at least be harmed, although some plants thrive in oxygen-deprived soils.

Normally soils with a mixture of noncapillary and capillary pores are preferred, because they both drain well and retain water for plant growth. When describing roots and pores, at least the abundance of each size class is assigned; for example, common fine roots or many medium pores.

Chemical Response

pH

The acidity or alkalinity can infer much information about a soil. The pH of a soil can be estimated with a simple field test kit. There are several soil pH class systems used to categorize soil pH. Each horizon or layer within a pedon is tested as pH may vary dramatically due to pedogenic or lithogenic properties.

Effervescence

A simple field technique may be used to check for the presence of free carbonates. A 10% solution of HCl added to the soil produces a gaseous response when

$CaCO_3$ is present. The gaseous response of the HCl will vary depending on the amount of carbonates in the soil. The degree of effervescence is noted in the description.

Reaction to Alpha,Alpha′-Dipyridyl

In soils that are actively reducing at an Eh (redox potential) or pE (relative electron activity) low enough to produce ferrous iron, a 0.2% concentration of alpha,alpha′-dipyridyl (also listed as 2,2′-bipyridine) may be used. This solution is applied to the soil within minutes of exposure in the absence of sunlight. If Fe^{2+} is present, the alpha,alpha′-dipyridyl will change to a red or pink color; if only Fe^{3+} is present, the solution will remain clear. This solution will react with a few parts per million of Fe^{2+}, so care should be used when extracting soil. The Fe^{2+} from a knife blade or spade contact with the soil will produce a false-positive response. This solution should only be used when reducing conditions are suspected.

Special Features

Pedon Features

Pedon features occur throughout the pedon and may cross adjacent horizons. These features are described using standard technical terms; however, some added language is sometimes used to supplement the description (Table 3).

Horizon or Layer Features

Concentrations are identifiable bodies within the soil that have been formed due to the pedogenic processes of chemical dissolution and/or precipitation, transport, accrual, and physical and/or biological removal. Concentrations may be finely disseminated materials, masses, nodules, concretions, crystals, biologic concentrations, and/or inherited minerals. The kind, size, abundance, and color are used when describing concentrations. In some soils, supplemental information is needed to describe concentrations fully.

Ped Surface Features

Surface features occur from a variety of pedogenic processes and result in differences from the matrix soil in composition, orientation, or packing. Surface features include: (1) coats of a variety of substances unlike the adjacent soil material and covering part or all of surfaces; (2) material concentrated on surfaces by the removal of other material; and (3) stress formations in which thin layers at the surfaces have undergone reorientation or packing by stress or shear. Descriptions of surface features may include kind, location, amount, continuity, distinctness, and thickness of the features. Where the features are thick and there are contrasting colors, this is noted. Table 4 lists some common ped surface features found in soils.

Location

Surface features may vary in the coverage of peds or may coat other soil constituents such as gravel or rocks.

Amount

The number should be noted as a percentage or a percentage range. Some horizons have more than one surface feature. In those cases, each surface feature is noted with the percentages of each feature.

Table 3 Common pedon features

Pedon feature	*Description*
Cracks	Fissures other than those attributed to soil structure
Desert pavement	Natural concentration of polished gravel and stones at the soil surface primarily in arid and semiarid climates
Ice crystals, lenses, layers (veins), wedges	Features occurring in soils of Arctic, Antarctic, and Subarctic regions
Jarosite mottles	Yellow jarosite mineral accumulations due to oxidation of sulfidic accumulations
Krotovinas	Irregular tubular streaks of soil material within a horizon or layer transported to another horizon or layer due to biologic activity
Lamellae	Thin banded illuviated clay
Microbiotic crust	Thin surficial zone exhibiting more mechanical continuity of the soil fabric than the subjacent soil
Stone line	Natural accumulation of coarse fragments below the surface extending through the pedon
Tongues of albic material	Lobes of eluviated soil that extend into the soil horizon from overlying horizons

Reproduced with permission from Schoenberger PJ, Wysocki DA, Benham EC, and Broderson WD (eds) (2002) *Field Book for Describing and Sampling Soils*, version 2.0. Natural Resources Conservation Service. Lincoln, NE: National Soil Survey Center.

Table 4 Common ped surface features

Ped feature	*Description*
Clay films	Layers of oriented, translocated clay ('clay skins')
Clay bridges	Links between mineral grains found primarily in sandy soils
Stress surfaces, pressure faces	Smooth, shiny or smeared surfaces formed through rearrangement of particles resulting from shear forces
Slickensides	Stress surfaces that are polished and striated; direction of movement can be determined from striations
Other features	Coats of iron, aluminum, or manganese oxides, organic matter, salts, or carbonates

Reproduced from Schoenberger PJ, Wysocki DA, Benham EC, and Broderson WD (eds) (2002) *Field Book for Describing and Sampling Soils*, version 2.0. Natural Resources Conservation Service. Lincoln, NE: National Soil Survey Center.

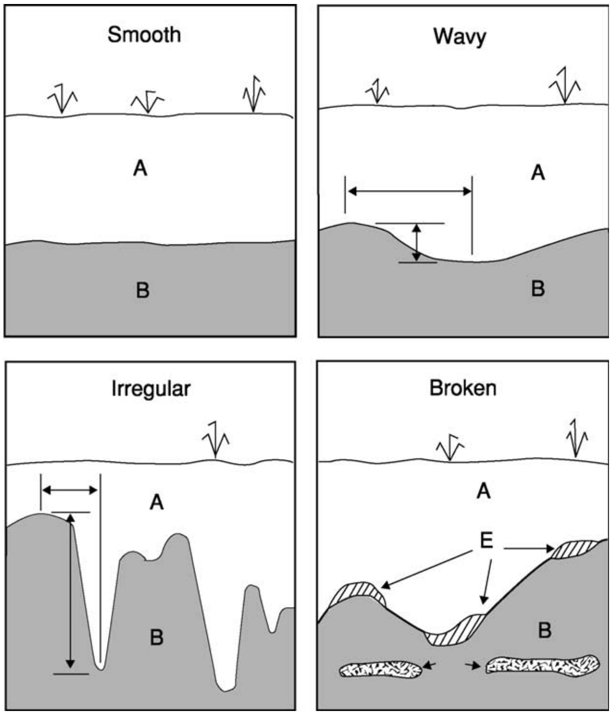


Figure 9 Boundary nomenclature of soil horizons. A, A horizon; B, B horizon; E, E horizon. (Reproduced from Schoenberger PJ, Wysocki DA, Benham EC, and Broderson WD (eds) (2002) *Field Book for Describing and Sampling Soils*, version 2.0. Natural Resources Conservation Service. Lincoln, NE: National Soil Survey Center.)

Boundary

The horizons or layers may be parallel to the surface; however, in some soils the bottom boundary may vary depending on the pedogenesis. In most description systems, the bottom boundary of the soil horizon or layer is described as smooth, wavy, irregular, or broken (Figure 9). The distance in which the boundary grades into or changes is normally described also. The boundary classes are abrupt (0–2 cm), clear (2–5 cm), gradual (5–15 cm), and diffuse (more than 15 cm).

Soil Micromorphological Description

The discussion in this chapter has focused on the macromorphological description of soils in the field. There is a discipline devoted to micromorphological descriptions in the laboratory using soil thin sections. These descriptions are valuable for classification and elucidating pedogenesis in soils. Where possible a micromorphologic investigation is used to supplement and verify features in the macromorphological description.

See also: **Texture**

Further Reading

Brewer R (1976) *Fabric and Mineral Analysis of Soils*. Huntington, NY: Kreiger.

Bullock P, Federoff N, Jongerius A, Stoops G, and Tursina T (1985) *Handbook for Thin Section Description*. Wolverhampton, UK: Waine Research.

Buol SW, Hole FD, McCraken RJ, and Southard RJ (1997) *Soil Genesis and Classification*, 3rd edn. Ames, IA: Iowa State University Press.

Childs CW (1981) Field tests for ferrous iron and ferric-organic complexes (on exchange sites or in water-soluble forms) in soils. *Australian Journal of Soil Research* 19: 175–180.

Schoenberger PJ, Wysocki DA, Benham EC, and Broderson WD (eds) (2002) *Field Book for Describing and Sampling Soils*. version 2.0. Natural Resources Conservation Service. Lincoln, NE: National Soil Survey Center.

Soil Survey Staff (1993) *Soil Survey Manual*. USDA, Soil Conservation Service, Agricultural Handbook No. 18. Washington, DC: US Government Printing Office.

Soil Survey Staff (1998) *Keys to Soil Taxonomy*, 8th edn. USDA, Natural Resources Conservation Service. Washington, DC: US Government Printing Office.

Vepraskas MJ (1992) *Redoximorphic Features for Identifying Aquic Conditions*. North Carolina Agriculture Research Service Technical Bulletin 301. Raleigh, NC: North Carolina State University.

Wilding LP, Smeck NE, and Hall GF (eds) *Pedogenesis and Soil Taxonomy. Concepts and Interactions*, part A. New York: Elsevier.

MULCHES

C L Acharya, K M Hati, and K K Bandyopadhyay,
Indian Institute of Soil Science, Bhopal,
Madhya Pradesh, India

Introduction

Mulch means a layer of dissimilar material separating the soil surface from the atmosphere, and mulching is the artificial application of mulch, practiced to obtain beneficial changes in the soil environment. Mulch may be organic (crop residues, stubble mulch etc.) or inorganic (plastic sheet, gravels etc.) in nature. It may be grown *in situ* like previous crop residue mulch left on the soil surface, produced by a cover crop or it may be living mulch like perennial legume mulch. It may also be grown or produced *ex situ* and brought in for field application, e.g., straw, sawdust, plastic products etc.

The practice of mulching has been widely used as a management tool for centuries in many ancient civilizations. The Great Plains region of the USA, which was once considered to be a dust bowl, has been turned into the granary of the world through the adoption of conservation tillage, the practice of leaving crop residues as the surface stubble or flattened straw mulch. The idea of using plastics as mulch originated in Hawaii, where it was used in pineapple cultivation with considerable success.

Mulch has a buffering effect and it dampens the influence of environmental factors on soil. The magnitude of the buffering effect of mulches depends upon the quality, quantity and durability of mulch material, soil type and the climatic conditions. Mulching influences the soil hydrothermal regime by influencing the radiation balance, rate of heat and water vapor transfer, and heat capacity of soil. Mulching improves the physical condition of soil by enhancing soil aggregation and helps in conservation of water by checking evaporation, increasing infiltration and retarding runoff loss. It favorably modifies the soil thermal regime, retards soil erosion, and improves soil chemical environment and biological activity in soil. Modification of the soil microclimate by mulching favors seedling emergence and root proliferation and suppresses weed population. Organic mulches add nutrients to soil when decomposed by microbes and help in carbon sequestration. Favorable soil edaphic environment under mulch improves crop productivity, enhances input-use efficiency and checks environmental pollution. From the conservation viewpoint, mulch simulates the effect of a plant cover. It is most useful as an alternative to cover crops in dry areas where insufficient rain prevents the establishment of a ground cover before the onset of heavy rain or strong winds, or where a cover crop competes for moisture with the main crop.

However, improper application of mulches leads to creation of an anaerobic environment under a heavy rainfall situation, leading to loss of nitrogen through denitrification. It may cause disease and pest infestation and some allelopathic materials produced in some crop residues retard the crop growth. So proper site-specific methodology needs to be followed under different soil, crop and climatic conditions for efficient mulch management to achieve sustainable crop production.

How Mulching Influences the Soil Environment

Mulching improves the physical conditions, chemical environment and biological activities of soil. Favorable modification of the soil hydrothermal regime, improvement of soil aggregation and retardation of erosion and soil loss enhance the physical condition of soil under mulch. Improvement in the soil chemical environment and biological activities by mulching is mainly attributed to the enrichment of soil organic matter through organic mulching and improvement of the physical condition of soil under both organic and inorganic mulches.

Soil Moisture Regime

Mulching favorably influences the soil moisture regime by controlling evaporation from the soil surface, improving infiltration and soil-moisture retention and facilitating condensation of water at night due to temperature reversals.

Controling evaporation from the soil surface Direct evaporation of water from the soil surface is an important process, particularly in the case of bare soils or in areas where summer fallow is practised. Mulching reduces evaporation from the soil surface by retarding the intensity of the radiation and wind velocity on the mulched surface. It is effective in retarding the evaporation mainly during the initial energy-limited stage of drying. However, mulching is not very effective in controling evaporation during the falling-rate or supply-controled stage of evaporation. During this stage, the soil surface dries under the mulch and the vapor transfer of water to the atmosphere is inhibited more by the dry soil than by the overlying vegetative mulch. Higher mulch rate retards the energy reaching the soil surface and hence limits the evaporation at the constant-rate stage (Figure 1). The retardation of initial evaporation can also enhance the process of internal drainage and thus allows more water to migrate downward into the deeper parts of the profile, where it is conserved longer and is less likely to be lost by evaporation.

Mulching also affects the pattern of evaporation by offering resistance to water vapor flow from the soil surface to atmosphere and by increasing the thickness of the relatively non-turbulent air zone above the soil, thus changing the boundary layer conditions of the soil–air interface. The relative magnitude of evaporation reduction for different mulch rates is much greater under high evaporative demand. The cumulative evaporation reduction due to mulching follows a quadratic relationship with time. The magnitude of maximum evaporation reduction and the time when it is reached depends on the rate of mulching, soil type, magnitude of evaporative demand and the method of residue application (Figure 2). The orientation of the residue, i.e., flat or matted versus standing, affects the porosity and thickness of the layer and thus the rate of evaporation from the soil surface. As the amount of standing residue increases, regter wind speed is needed to initiate water loss through evaporation; in addition, the water loss rate at a given wind speed decreases with increased amounts of standing residue. Residue position also affects soil temperature (Table 1), which in turn affects the evaporation rate through its influence on the vapor pressure of the soil water.

Water-saving under mulching is prominent if rains are frequent. However, under extended dry spell periods, mulch may keep the soil surface moist for longer periods and hence prolong the first stage of evaporation and thus produce no net saving of water.

Improving infiltration rate Mulching with organic materials improves the infiltration rate because it serves as a barrier for runoff, which provides more opportunity for water to infiltrate into the profile. Secondly, mulch intercepts the rainwater, protects the soil from its biting action and checks splashing of the soil. It prevents crust formation due to clogging of soil pores, which increases infiltration rate. Furthermore, organic mulches improve the macroporosity and stability of the structural aggregates of soil

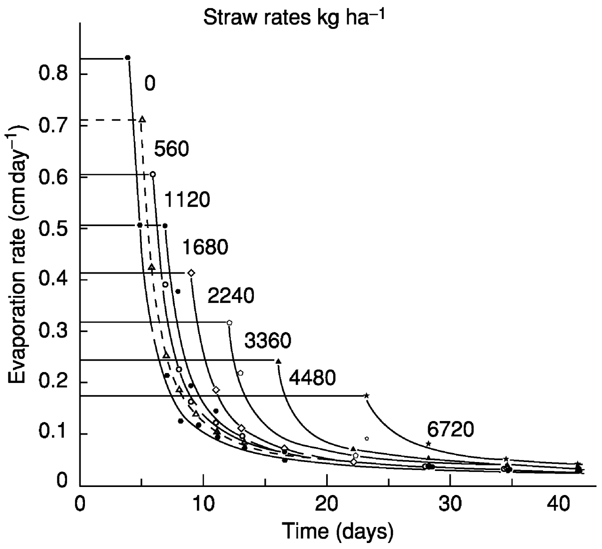


Figure 1 Temporal variation of soil water evaporation rate as influenced by rates of straw mulch application. Straw rates range from 0–6720 kg ha^{-1}. Reproduced with permission from Bond JJ and Willis WD (1969) *Soil Science of America Proceedings* 33: 446.

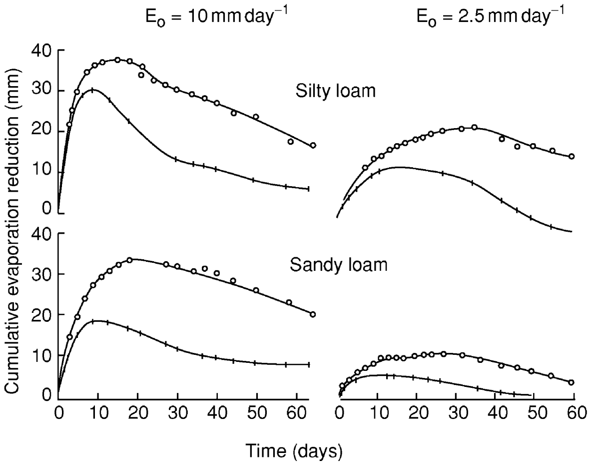


Figure 2 Time-course of cumulative evaporation reduction (CER) in two different soils under two rates of wheat straw applied as mulch at two evaporative demands (E_o). Wheat straw. ⊢—⊣ 3 t ha^{-1} as mulch; ○—○ 6 t ha^{-1} as mulch. Reproduced with permission from Prihar SS, Jalota SK, and Steiner JL (1996) *Soil Use and Management* 12: 152.

Table 1 Average daily soil surface temperature as affected by wheat straw position during August to September at Colorado, USA

Straw position	*Soil surface temperature (°C)*[a]
Bare soil	47.8 c
Flat straw	41.7 b
3/4 flat and 1/4 standing	39.6 b
1/2 flat and 1/2 standing	32.2 a

[a] Column values accompanied by different letters are significantly different at $P = 0.01$ (Duncan's multiple range test).

Reproduced from Smika DE (1983) In: Unger PW (1990) *Advances in Soil Science* 13: 27–68.

and thereby improve the water transmission properties (Table 2), which facilitate better infiltration and recharge of the soil profile. Straw mulch application increases soil water storage and storage efficiency. The amount of water storage in the soil profile, the storage efficiency, total water use and water-use efficiencies of dryland crops increase with increase in the mulch rate (Table 3). The increase in infiltration that results from mulch is found to be more important in some situations than its effect on reduction in evaporation for conservation of water in the profile.

Soil moisture retention Mulching improves the moisture retention properties of soil through its effect on pore-size distribution and soil structure. Higher mulch rates increase soil-water retention more at lower suctions (Table 4) due to an increase in macropores and interaggregate pores caused by enhanced soil organic matter content and higher activity of soil fauna, e.g., earthworms and termites in mulched plots.

Table 2 Effect of mulch rate on water transmission properties of an Alfisol

Mulch rate (Mg ha^{-1})	*Soil water sorptivity*	*Soil water transmissivity*	*Saturated hydraulic conductivity (cm h^{-1})*[a]
0	0.32	5.56	30a
2	0.57	7.81	45a
4	0.67	7.50	70b
6	0.84	10.21	132c
12	1.05	15.36	129c

[a] Column values followed by the same letter are not significantly different at the 5% level (Duncan's multiple range test).

Reproduced from Lal R, De Vleeschauwer D, and Malafa Nangje R (1980) Changes in the newly cleared Tropical Alfisol as affected by mulching. *Soil Science Society of America* 44: 827–833.

Table 3 Effect of straw mulch rate on soil water storage during fallow,[a] water storage efficiency and dryland grain sorghum yield in Bushland, Texas

Mulch rate (Mg ha^{-1})	*Water storage (mm)*	*Storage efficiency (%)*[b]	*Grain yield (Mg ha^{-1})*	*Total water use (mm)*	*Water use efficiency (kg m^{-3})*
0	72 c	22.6 c	1.78 c	320	0.56
1	99 b	31.1 b	2.41 b	330	0.73
2	100 b	31.4 b	2.60 b	353	0.74
4	116 b	36.5 b	2.98 b	357	0.84
8	139 a	43.7 a	3.68 a	365	1.01
12	147 a	46.2 a	3.99 a	347	1.15

[a] Fallow duration 10 to 11 months.

[b] Precipitation averaged 318 mm.

Column values followed by the same letter are not significantly different at 5% level (Duncan's multiple range test).

Reproduced from Unger PW (1990) *Advances in Soil Science*, 13: 27–68.

Water condensation at night Stone and gravel mulches induce lateral movement of heat and vapor, which could in turn collect water under the stones, due to condensation of vapor at night, in amounts sufficient enough to serve as the source of water for some species of desert plants.

Controlling Runoff and Soil Erosion

Mulching invariably decreases soil erosion and often reduces the rate and amount of runoff. Mulch cover protects the soil from raindrop impact and surface sealing, increases the infiltration rate and decreases runoff velocity through physical resistance to water flow. In general, loss of water through runoff decreases exponentially with increase in mulch rate (Table 5). For some soils, mulch does not substantially decrease runoff and drastically reduces soil erosion. The runoff water gets filtered through the mulch and is often clear with little sediment. Mulching decreases sediment concentration in runoff water through the protective effect of crop residues against raindrop.

The rate of soil loss decreases exponentially with the increase in the percentage area covered by mulch. The mulch factor (MF), defined as the ratio of soil loss with mulch to that without mulch, is related to the percentage residue cover (RC) or mulch cover by the following expression:

Table 4 Effect of mulch rate on moisture retention properties of 0–5 cm layer of an Alfisol

	Moisture retention (% w/w) at different suction (bar)			
Mulch rate (Mg ha^{-1})	0.03	0.06	1	15
0	27.3	26.1	5.5	3.2
2	29.1	26.1	6.5	5.1
4	29.6	28.8	5.2	3.7
6	29.6	28.5	5.4	4.2
8	31.2	30.5	7.3	5.8

Reproduced from Lal R (1987) *Tropical Ecology and Physical Edaphology*, pp. 618–685. Wiley-Interscience Publication.

Table 5 Effects of rice straw mulch rates on runoff and soil loss in Nigeria

Mulch rate (Mg ha^{-1})	*Runoff (%)*	*Soil loss (Mg ha^{-1})*	*Marginal conservation effect (Mg soil) (Mg mulch)$^{-1}$*
0	75.4	9.6	–
2	43.4	2.3	3.7
3	15.2	0.5	1.8
6	5.4	0.1	0.13
12	0	0	0.017

Reproduced from De Vleeschauwer *et al.* (1980) In: Erenstein O (2002) *Soil and Tillage Research* 67: 115–133.

$$\mathrm{MF} = \mathrm{e}^{-\mathrm{aRC}}$$

where, ‘a’ ranges in value from 0.03 to 0.07 and 0.05 is generally used. The exponential relationship is applied only for rill erosion whereas the interrill erosion decreases linearly with the mulch cover. Since for most soils the contribution of the interrill erosion to the total soil loss is quite small for slope length in excess of 25 m and for all but very low percentage residue covers, the above equation may be used to compute the mulch factor. This mulch factor is multiplied with the ‘C’ factor values for calculating the effect of mulching in the universal soil loss equation. It has been observed that a minimum of 50% soil cover is required to achieve reduction in soil loss. Ideally a mulch should cover 70–75% of the soil surface. The optimum mulch rate can be computed for a pre-selected set of conditions using the Manning’s equation for flow velocity. The relationship between Manning’s n and mulch rate is given below:

$$\text{Manning's equation:} \quad n = (\mathrm{r}^{0.67} \times \mathrm{S}^{0.5})/\mathrm{v}$$

where, r = hydraulic radius, S = slope and v = desired maximum flow velocity.

The value of n due to mulch has been described as $n_\mathrm{m} = (n^{3/2} - n_\mathrm{s}^{3/2})^{2/3}$ where, n_s = Manning’s n for a particular soil.

Mulch rate (M, kg m^{-2}) is related to the Manning’s n for mulch (n_m) by the following relationship.

$$n_\mathrm{m} = 0.071\,\mathrm{M}^{1.12} \quad \text{for interrill erosion}$$

$$n_\mathrm{m} = 0.105\,\mathrm{M}^{0.84} \quad \text{for rill erosion.}$$

Experimental evidence shows that an adequate quantity of crop residue mulch can mitigate the effect of the degree of slope in reducing soil loss.

Mulching is also effective in reducing soil erosion by wind. It reduces the wind velocity on the soil surface when the mulch is fixed to the soil, so that soil particles do not get blown up by wind. Standing crop residues are more effective than flattened surface mulches in reducing wind erosion. The effectiveness of standing crop residues depends upon the number of stalks per unit area and their size.

Modifying Soil Thermal Regime

Mulch has a moderating influence on the soil thermal regime and the effect varies among soils, climate, kind of mulch materials used and rates of application. It increases soil temperature during cooler weather and decreases it during hot spells. In general, mulch has a damping effect on the amplitude of the diurnal fluctuations in soil temperature. Organic mulching enhances the soil temperature at night and early morning hours but it decreases the daytime temperature as compared to unmulched plots (Figure 3).

Mulching is commonly practiced in some fruit crops, like strawberry, to keep the fruit clean by avoiding direct contact with soil and to moderate the soil hydrothermal regime for better productivity and quality of fruit. Transparent polythene mulch raises the maximum soil temperature whereas organic mulches like pine needle and grass mulches lower it. Black polythene mulch, however, does not alter the maximum soil temperature appreciably (Figure 4). The application of straw mulch can lower maximum soil temperature due to the interception of incoming solar radiation, high reflectivity and low heat conductivity, the magnitude of which depends upon soil wetness, incidental radiation, rate of mulch application as well as period of the year (Table 6). The lowering of maximum soil temperature by straw mulching during the early stage of growth of sugarcane can significantly improve the yield.

Soil Aeration

Crop residue mulch improves soil aeration by promoting free exchange of gases between the soil and the atmosphere. This is facilitated by improvement of structural stability, total porosity and macroporosity, decrease of surface crusting and by improving the overall soil drainage. Oxygen diffusion rate is higher under mulch than under unmulched conditions. The gaseous composition of soil air under mulch depends on the nature of the mulch material (C:N ratio), its rate of decomposition, the soil moisture regime and the climatic condition. Plastic mulch is practically impervious to carbon dioxide (CO_2), a gas that is of prime importance for photosynthesis. Very high levels of CO_2 build up under the plastic, as the film does not allow it to escape. It has to come through the holes made in the plastic for the plants and a ‘chimney effect’ is created, resulting in localized concentrations of abundant CO_2 for the actively growing leaves that accelerates the growth of the crop.

Improving Soil Structure

Mulching improves soil structural properties directly by preventing raindrop impact and indirectly by promoting biological activity. Organic mulching improves the total porosity, macroporosity and mean weight diameter of water-stable aggregates, because of addition of organic matter upon decomposition by soil microorganisms. The mean-weight diameter of water-stable aggregates increases with increase in the mulch rate (Table 7).

In general, under mulched conditions, the bulk density of the soil is lower than that under unmulched

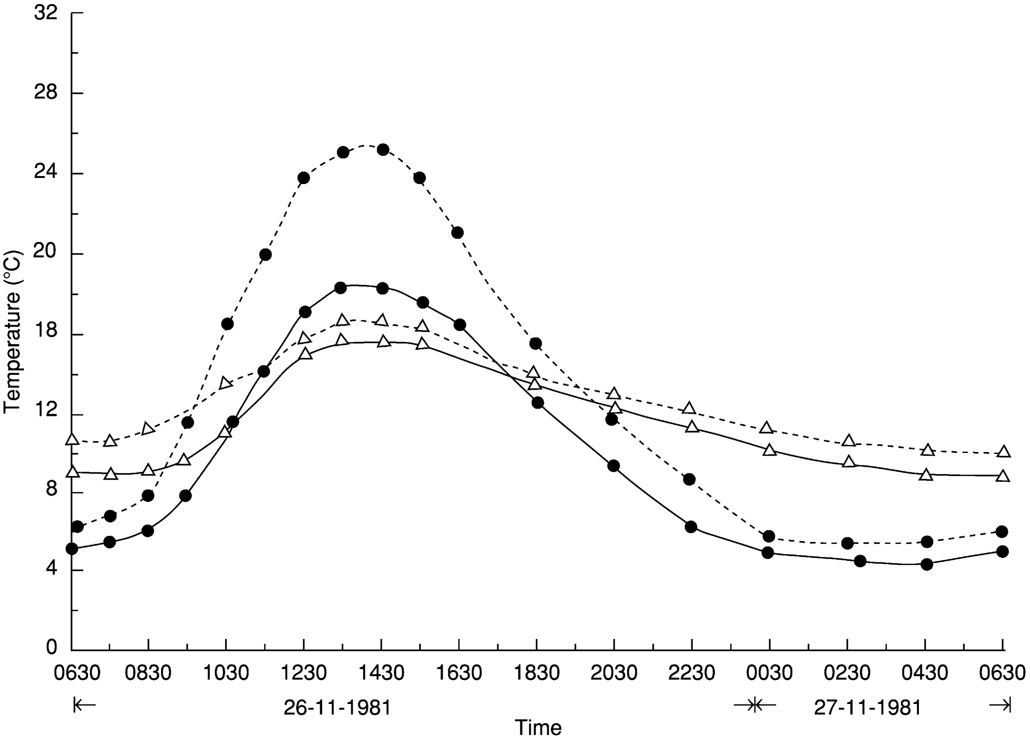


Figure 3 Diurnal soil temperature fluctuations as influenced by conventional cultivation (CC); and pine needle mulching (M) at 10 Mg ha^{-1} in wheat on a silty-clay loam soil. ● CC; △ CC+M; — 10 cm depth; - - - 5 cm depth. Reproduced from Bhagat RM and Acharya CL (1987) *Soil and Tillage Research* 9:65–77, with permission from Elsevier Science.

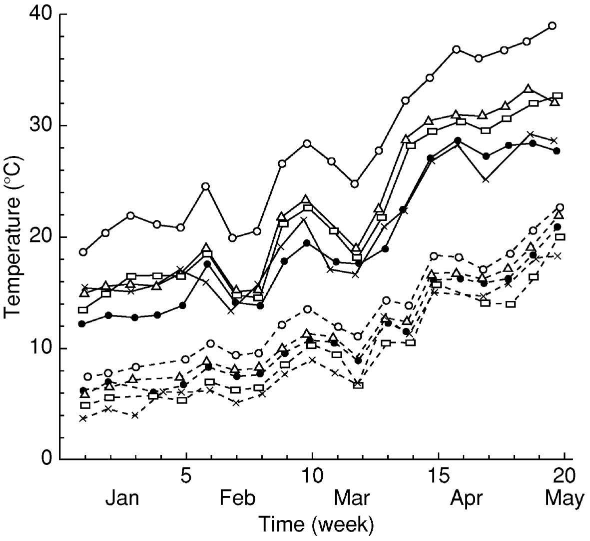


Figure 4 Mean weekly soil temperature at 5 cm depth under different mulch treatments and unmulched control in strawberry on a Typic Hapludalf. ○, transparent polythene; △, black polythene; □, unmulched control; ●, pine needle mulch; — temperature at 14.30 hours; - - - temperature at 06.30 hours.

conditions. Bulk density decreases with increase in residue mulch rate. One of the reasons for the decrease in bulk density with increase in mulch rates is high earthworm activity in mulched plots.

Table 6 Effect of straw mulching (@ 4–6 Mg ha^{-1}) on soil temperature at 10 cm depth during the year under different crops

Month	*Crop*	*Maximum temperature of unmulched soil (°C)*	*Temperature reduction with straw mulch (°C)*
January	Autumn potato	12.5–15.5	0.5–2.0
March	Spring potato	24.5–29.0	2.5–3.8
April	Sugar cane	37.6–38.2	8.1–9.5
	Bare soil	38.3[a]	11.5
May	Fodder maize	38.0–42.5	7.5–9.8
	Bare soil	40.3[a]	6.7
June	Sugar cane	34.2–43.5	4.4
July	Grain maize	35.5–41.7	3.7–6.8
August	Grain maize	31.7–38.9	1.4–4.8
September	Grain maize	31.5–33.2	0.9–2.0
November	Autumn potato	17.0–18.5	2.0
December	Autumn potato	15.2–17.0	1.7–2.2

[a] Recorded at 5 cm soil depth.

Reproduced from Prihar SS (1986) In: *Proceedings of Seminar Organized by Fertilizer Association of India* (Northern Region) at Dehradun, India, May 10–11, 1986, pp. 31–44.

Improving the Soil Chemical Environment

Organic mulches add organic matter and plant nutrients to soil upon decomposition. Thus, they improve carbon sequestration. The volatilization and leaching

Table 7 Effect of mulch rate on pore-size distribution of 0–5 cm layer of an Alfisol after 18 months of experiment

	Porosity (%)			
Mulch rate (Mg ha^{-1})	*Macro*	*Meso*	*Micro*	*Mean weight diameter (mm)*
0	17.9	7.1	23.1	1.2
2	20.6	7.2	22.4	1.4
4	27.7	8.4	18.7	2.1
6	28.9	8.3	17.9	2.0
12	37.0	8.2	13.2	2.4
LSD (0.05)	5.3	NS	4.2	–

Reproduced from Lal R (1987) *Tropical Ecology and Physical Edaphology*, pp. 618–685. Wiley-Interscience Publication.

loss of nitrogen is reduced under mulched conditions. During decomposition of organic mulches, soil mineral nitrogen is immobilized by microbes and thus its loss is minimized. Cation exchange capacity is substantially influenced by organic matter content in soils containing predominantly low-activity clays. Improvement in the cation exchange capacity of soil improves the fertility status of these soils. Furthermore, decomposition of organic mulches adds organic acids to the soil resulting in low soil pH, which influences the bioavailability of many plant nutrients, e.g., Fe, Mn, Zn, Cu, etc. Mulches also influence the availability of nutrients through their influence on physical conditions, hydrothermal regime and biological activity of the soil.

Improving Biological Activity in Soil

Soil biological activity is either directly influenced by the supply of food substrates by organic mulches or indirectly influenced by both organic and inorganic mulches through alteration of the soil hydrothermal regime. The activity and diversity of soil organisms is substantially influenced by the quality of mulch material and the rate of its application. The soil microorganisms (bacteria, fungi and actinomycetes) and soil fauna (earthworm, millipede, centipede etc.) help in the decomposition of soil organic matter and hence the availability of nutrients. However, mulching with materials of high C:N ratio results in immobilization of nitrogen by microorganisms. Mulching with organic residues of weeds like wild sage (*Lantana camara*) or eupatorium (*Eupatorium adenophorum*) under conservation tillage practices enhances the earthworm population due to supply of food substrate and a better hydrothermal regime under mulch (Table 8).

Table 8 Earthworm population (number m^{-2}) in the surface (0–15 cm) layer at sowing and flowering stages of wheat

	Number of earthworms	
Treatments	*Sowing*	*Flowering*
Mulch tillage (T)		
T_1: Mulching previous standing maize with wild sage followed by its incorporation with conventional tillage immediately prior to sowing of wheat	20	8
T_2: As T_1 but wheat sowing following conservation tillage and retaining the residues of wild sage as mulch	21	10
T_3: As T_1 but mulching by eupatorium instead of wild sage	21	8
T_4: As T_2 but mulching with eupatorium instead of wild sage	20	10
T_5: Farmer's practice of conventional tillage after maize harvesting	4	1
LSD ($P = 0.05$)	1.7	1.4

Adapted from Acharya CL, Kapur OC, and Dixit SP (1998) *Soil and Tillage Research* 46: 153–163.

Types of Mulches

There are different types of materials ranging from natural plant residues, to gravel to various industrial products such as paper or plastic sheets, which are used as mulch. These can broadly be categorized into two groups i.e., organic mulch and inorganic mulch.

Organic Mulches

Organic mulches include living mulch, cover crops and mulches whose materials are derived from natural sources, such as field crops, trees, or natural waste products. These include hay, straw, wood chips, bark, sawdust, wood shavings, peat moss, compost, grass clippings, paper, and materials such as cocoa shells and buckwheat hulls. Organic materials for mulching should be a stable material that will remain in place at least through the growing season of an annual crop and longer for perennial crops. More durable organic material provides longer-lasting benefits as mulches than readily decomposable plant residues. Durability of organic mulch material depends upon its composition. Factors influencing the rapidity of residue decomposition include the nature of the material like the C:N ratio, cellulose and lignin content, climatic conditions, and types and population of soil microorganisms. The decay rate of surface-applied plant residues indicates that the high lignin and cellulose-containing cereal straws produce more weather-resistant mulch than do legumes.

The effectiveness of organic or vegetative mulches also depends upon the thickness of the mulch. A lesser covering does not adequately protect the soil while a greater covering suppresses plant growth. Vegetative mulches, due to their low thermal conductivities, decrease the conduction of heat into and out of the soil, causing the soil underneath to be cooler during the day and warmer during the night than bare soil. Thus it acts as an insulating agent on the soil surface. Vegetative mulches significantly affect soil temperature through their effect on the radiant energy balance of the system. The radiation balance is influenced by reflection of incoming radiation by surface residues, differential heating of soil and air and alteration of evaporation of soil water from the surface. Soil temperature generally decreases as reflectance increases. The insulating effect of mulches improves with increase in thickness of the surface residue layer. Age of the residue, color, geometry (standing, flattened or matted), distribution and amount are the main characteristics that affect radiation reflection. Reflection is normally greatest from bright residues like wheat straw and decreases with residue aging (discoloration) and decomposition. As a result, temperature differences between residue covered and bare soils are greatest when residues are bright and decrease as they age and decompose.

Organic mulches have the advantage of being biodegradable. In time, they are broken down and consumed by microorganisms, adding to the soil's organic matter content and improving soil structure or tilth. The vegetative mulches reduce the rate of evaporation from the soil surface, but the effectiveness of vegetative mulching in the reduction of evaporation depends upon the thickness, type, orientation and color of the residue. The rate of evaporation decreases with an increase in mulch thickness. As the material density of mulch affects the thickness and surface coverage obtained with a given weight of material, a low-density material such as wheat straw reduces evaporation more effectively than higher-density materials such as sorghum stubble or cotton stalks. To obtain a reduction in evaporation similar to that obtained with wheat straw on a weight basis, about twice as much sorghum stubble and more than four times as much cotton stalks are needed.

Living or cover crop mulch is also an important organic mulch used for some specific crops. Living mulch is a covering of living plant material, which is planted or allowed to remain living on the land after another crop has been planted. The live mulch system is based on the principles of mixed cropping. Here a perennial legume is established with the objective of smothering weeds and a seasonal grain crop is grown in small strips through the established live cover crop. The system works if the live mulch is a low-growing, nonclimber and does not compete for light, moisture or nutrients with the main crops. The live mulch system is more successful in humid and sub-humid regions with little or no water deficit than in semiarid or arid regions. This system imparts some of the benefits similar to surface-applied organic mulches such as increased soil carbon content, increased rainwater infiltration and reduced soil erosion. Cover crop mulches suppress weed growth through modification of the microclimate and release of allelochemicals. By land-surface coverage, mulches block the light stimulus required for germination of many small-seeded weed species. However, drastic yield suppression of the food crops has been reported in some studies due to allelopathic effects of the cover crop on the main crop and competition for moisture during periods of drought stress. Some important crops like cereal rye (*Secale cereale* L.) and hairy vetch (*Vicia villosa* Roth) can be used as cover crop mulch. Cereal rye is one of the best winter-hardy cover crops grown in the mid-western USA. Perennial ryegrass (*Lolium perenne* L.), a potential cover crop, is commonly used as a living mulch for some vegetables like cabbage.

Paper mulch is an important organic mulch produced from recycled paper. It has been evaluated as an available biodegradable alternative to plastic mulch. Paper does not require disposal at the end of the season and the use of recycled paper can reduce solid waste as well as energy costs. But paper may decompose too rapidly to maintain season long ground coverage. This can be checked by treating the paper with waste cooking oil, which delays the breakdown and enhances soil warming by increasing the paper's translucency to sunlight.

Inorganic Mulches

Different inorganic materials such as plastic film, gravels, crushed rocks etc. are used as mulch. Inorganic materials used for mulches do not add nutrients or humus to soil and do not decompose except after long exposure to weathering. Otherwise these materials are effective mulches and several of them are permanent and quite attractive.

Plastic mulch Plastics are the most widely used inorganic mulch materials. Mulching with the help of plastic film has played a major role in plant cultivation by creating some measure of mechanical protection at the soil surface and a microclimate favorable from the aspects of temperature distribution, retention of humidity and the supply of carbon dioxide to the stomata of lower leaves of small plants. Plastic mulch has been used commercially for vegetable production since the early 1960s. It can be used very effectively with trickle or drip irrigation, particularly for the cultivation of vines and strawberries where it encourages the lateral development of the root system.

In many of the developed countries plastic film, such as low-density polyethylene (LDPE) film is rapidly replacing conventional vegetative mulching because of its effectiveness and commercial viability, particularly in the case of high-value crops. In this method around 30–60% of the cultivated area is covered with a thick polythene film (30–150 μm). The plants are allowed to grow through the holes made in the LDPE film. In ornamental plants it is often used under gravel or stone mulches. The effectiveness of the plastic mulch depends upon the color of the mulch. Color determines the energy-radiating behavior of the mulch and affects the surface temperature of the mulch and the underlying soil temperature.

Recent new innovations in plastic technology manufacturing have resulted in the production of new plastic products for the farmers, e.g., photodegradable mulch, co-extruded mulch (black/white), and different colored mulches that serve specific purposes like maintenance of cool soil temperature during summer in tropical countries, increased warming of soil in spring in temperate regions and repelling aphids with silver-colored mulch. By raising the soil temperature in the planting bed, plant growth is accelerated, producing earlier yields in the temperate region. Black plastic mulch can result in harvest 7–14 days earlier while clear plastic accelerates harvest by 21 days in some conditions. For enhancing the earliness of vegetable crops in temperate regions, the mulch should maximize the transmission of the soil-warming portion of sunlight (the near infra-red radiation) and minimize the amount of visible light transmitted, as this visible light increases the weed growth under the mulch.

Due to the high cost of the cover and the difficulty of keeping it in place, plastic mulches are mainly used for some high-value vegetable and cash crops where these are found to be cost-effective. Muskmelons, tomatoes, peppers, cucumbers, squash, eggplants, watermelons and okra are some of the vegetable crops that have shown significant increases in earliness, yield, and fruit quality when grown on plastic mulch. Some less-valuable crops like sweetcorn, snap beans, southern peas and pumpkins have also shown similar responses in temperate climates. Based on the ability to transmit, absorb or reflect specific wavelengths, plastic mulches are divided into the following categories.

Black plastic mulch: Black is the most commonly used color of polyethylene mulch. This is chosen more often than clear polythene mulch because of its ability to control weeds by excluding photosynthetically active radiation. Black polythene placed in direct contact with the soil can warm the soil down to a depth of about 8–10 cm. Under black plastic film the temperature of the soil during the day is not significantly higher and in certain cases it is even lower than in the bare ground but during the night the temperature is always higher by 0.5–4°C than uncovered ground. This is desirable for warm-season vegetables such as peppers, tomatoes, eggplants etc. It is also widely used for the cultivation of pineapples in Hawaii, USA for reaping a higher yield.

Transparent/clear plastic mulch: Clear plastics have the greatest warming potential. They are transparent to incoming radiation and trap the longer wavelengths re-radiating from the soil, thereby producing a greenhouse effect that warms the soil underneath. Under transparent plastic film during daytime, soil temperature remains 2–10°C more than the bare ground while at night a difference of 2–4°C is observed, depending on the season, soil type and its moisture content. For warm-season crops, such as vine crops, earlier and higher yields can be achieved with clear mulch. However, as clear mulch is transparent to visible light, weed infestation under this mulch poses some serious problems, which may negate the benefits of soil warming.

White plastic mulch: Under white plastic mulch the temperature always remains lower than the uncovered ground due to its high reflectance and low transmittance of solar radiation. Since energy exchange is very poor, this film is used either in regions with a high level of sunshine where it is required to reduce the energy and lower the soil temperature or in regions of low luminosity where the amount of reflected light is to be increased on the lower and middle leaves of crop plants.

Colored plastic mulch: These are opaque, like black plastics, and are made with many different pigments. They do not allow light to pass through, but rather, reflect specific wavelengths of light up into the crop canopy. These mulches may be useful for enhancing crop production. For example, red mulches are effective in promoting growth of tomatoes. The other response observed with different mulch colors is the increase or decrease in insect population depending on reflected light and possibly temperature. Yellow appears to attract insects, especially cucumber beetle, and silver appears to repel aphids. The possibility of utilizing yellow mulch to monitor insect populations in a field of non-yellow plastic mulch shows promise. In addition, reduction of pesticide applications could be possible if only rows of yellow mulch (one row of yellow for every five or six rows of other-colored mulch laid) were sprayed with insecticide to control high populations of pests. Thus, yellow plastic mulches can be used to attract insects, much like a trap crop.

Infrared transmitting mulches (IRT): These have been developed recently and are wavelength-selective mulches. IRT mulches warm the soil intermediate to clear and black polythene mulches. These mulches are pigmented to reduce the amount of visible light transmitted and thereby reduce weed growth underneath. Thus, they blend the soil-warming characteristics of clear mulch with the weed-control ability of black mulch. IRTs contain very specific pigments that give them the unique ability to transmit a maximum of near-infrared radiation and a minimum (14–16%) of visible light. Some visible light is required to be transmitted in order to maximize soil warming. The benefits of IRTs are greatest for early-season plantings in temperate regions. Though these mulches are more expensive than black or clear plastics, some growers find them useful for high-value, heat-loving crops like melons.

Gravel mulch Gravel mulching is an age-old practice and can be very effective in water conservation (both in enhancing infiltration and in suppressing evaporation, even in layers as thin as 5–10 mm). Pea gravel is an attractive permanent mulch that is usually applied 5–10 cm deep and can be reused indefinitely. These are particularly good for rock garden plantings. The problem of this type of mulches is that once applied, gravels cannot be removed, do not decompose and might hinder future farm operations or land uses. Another issue is its high cost of mining, transportation and application in the field.

Crushed rock Crushed volcanic rocks or stones are available in many colors or sizes and make a permanent cover. These materials are especially useful around plants subject to crown rot. The color of the rocks influences the temperature of the soil underneath. Black rock absorbs heat and keeps the soil temperature warmer than normal, while white rock reflects sunlight and thus maintains a cooler temperature underneath. However, white rock can generate too much heat for the above-ground portion of some plants to survive.

Other materials used as mulch Sprayable preparations of latex, asphalt, oil, fatty alcohols etc. are also used as mulch for certain high-income crops such as strawberries. They inhibit evaporation by forming monomolecular films on the soil surface. Some water-proofing chemicals like silicone on application to clods on the soil surface form a layer of water-repellent aggregates. These aggregates then constitute a dry mulch that serves as an evaporation barrier. This layer also enhances infiltration by allowing the treated surface zone to act as a detention reservoir to prevent runoff.

Methods of Mulch Application

Proper methodology and timing of mulch application should be followed to realize the full benefit of mulching. The methods of mulch application vary with soil type, climate, crop and type of mulch material used.

Paper and plastic mulches are generally spread and fastened down either between the rows or over the rows of the crop. The plant seedlings in the latter case are allowed to grow through suitable openings made in the mulch material. Plastic mulches in commercial crop cultivation are generally applied by mulch applicators, which are available commercially or can be manufactured. They generally consist of a plastic-roll holder, two furrow-openers, and two land wheels followed by two furrow-closers mounted on a tool bar. The furrow-openers create a trench on either side of the mulch. The two land wheels follow the furrow-openers and hold the plastic in the furrow. The two furrow-closers refill the trench, firmly anchoring the plastic. A 15–23 cm border of plastic is usually buried leaving 76–91 cm of mulch exposed. Rototilling prior to installation creates a smooth lump-free surface, which helps to ensure a tight union between the plastic and the soil for more efficient heat transfer.

Organic mulches are normally spread a few centimeters thick on the soil surface. The depth of mulch depends upon the mulch material and types of crop on which they are applied. Thicker layers of mulch are normally placed around trees and shrubs, while finer and smaller materials are generally used for small flowers and vegetables. Ten centimeters of loose fibrous materials work well around trees and shrubs. The finer and smaller the particle size, the thinner the layer needs to be. Thick layers of very fine material block air to the roots of plants. For trees, mulch cover should reach at least to the tree's drip line or the outer perimeter of the branches. It should preferably be kept at least 15 cm away from the tree trunk and a couple of centimeters away from plant stems.

Mulches are normally applied and spread uniformly on the soil surface except in the case of vertical mulching. In vertical mulching crop residues, stubbles of the previous crop or other agricultural waste materials are dumped in vertical trenches dug at few-meter intervals in the field with the main objective to recharge the profile and conserve moisture in low rainfall areas of the semiarid tropics. The few centimeters-protruded portions of the stubble from the trenches also reduce the runoff and sediment loss from the field.

Impact of Mulching on Crop Growth and Input-Use Efficiency

Modification of Crop Growth

The effect of mulches on crop growth response depends on the combination of climate, type of crop and soil. Mulching influences three principal factors affecting crop growth, i.e., weed competition, seedling emergence and root growth, which in turn influence crop productivity.

Weed competition Straw mulching suppresses weed growth through delayed emergence, reduced weed count and smothering effects. Effectiveness of weed control is dependent on the type of mulch material used. Black plastic mulches are more effective for weed control than organic mulches in thin layers as light does not penetrate through black plastic mulches.

Seedling emergence Lack of moisture or unfavorable soil temperature in the seed zone often limits seedling emergence even though there may be sufficient water in the lower part of the profile. Mulch keeps the surface layers wetter for longer periods and thus enhances seed germination and seedling emergence. The times required for initiation of emergence and maximum emergence are shorter under mulching than under no mulch. Straw mulch significantly increases percentage of seedling emergence both under dry spell and rainfall situations (Table 9). Under dry-spell conditions, seedling emergence increases with mulching due to lowering of maximum temperature and reduction in evaporation loss of water. Under rainfall conditions the seedling emergence increases under mulching due to retardation of crust formation, which otherwise hampers seedling emergence.

Table 9 Effect of wheat straw mulching on seedling emergence in cotton and soybean in a sandy loam soil

	Percent seedling emergence			
	Cotton		*Soybean*	
Treatment	*No rain*	*Simulated rain*	*No rain*	*Simulated rain*
No mulch	25	9	19	0
Mulch	31	23	30	20

Adapted from Prihar SS (1986) In: *Proceedings of Seminar Organized by Fertilizer Association of India* (Northern Region) at Dehradun, India, May 10–11, 1986, pp. 31–44.

Root growth Mulching favorably influences root growth and distribution in the upper soil layer due to favorable soil temperature and higher soil water content under mulch. Higher soil water content in the surface layer reduces the mechanical resistance to growing roots, which helps in their proliferation. Application of organic mulches to the previous standing crop after the recession of rains helps in conservation and carry-over of higher amounts of residual soil moisture at the time of sowing of the subsequent rainfed crop, leading to better root proliferation and higher crop yield in regions where the weather starts cooling in the post-rainy season (Table 10). This carry-over effect is more prominent under conservation tillage than under conventional tillage.

Improvement of Input-Use Efficiencies

The favorable effects of mulching on crop growth and yields through an improved soil environment enhance both water and nutrient-use efficiency. Increase in water availability in the upper layers due to a reduction in evaporation and runoff losses under mulch is one of the reasons for increased input-use efficiency. Furthermore, modification of the soil hydrothermal regime under mulch favors root proliferation, which can tap water and nutrients from a larger soil volume, thus improving water and nutrient-use efficiency. Mulching results in better coordination of demand and supply of water resulting in complete extraction of profile-stored water, which improves the water-use efficiency. As grain yield-water use relation is linear within the domain of limited water supplies, any increase in water supply, beyond a minimum threshold would be directly reflected in increased grain yield as well as water-use efficiency (WUE).

Higher soil water content particularly in the initial stages of crop growth results in greater total water use by the crop, leading to higher plant growth and greater yield. Another factor contributing significantly to higher yield is more efficient use of growing season precipitation when the surface is covered with mulch. The final consequence of a higher yield with a relatively small increase in total water use is higher WUE with mulching than under unmulched conditions.

Application of pine needles, a waste product from pine forests, as mulch at the time of sowing of some tuber crops like potato improves the yield and water-use efficiency and results in savings of nitrogen fertilizer and irrigation water (Table 11). Hence, within certain limits, mulches can act as a substitute for both irrigation water and nitrogen fertilizer.

Table 10 Effect of tillage and nitrogen on root growth and grain yield of rainfed wheat in a silty-clay loam soil (Typic Hapludalf)

		Root mass density (kg m^{-3})		
Mulch tillage (T)	*Nitrogen*	*Tillering*	*Flowering*	*Grain yield (Mg ha^{-1})*
T_1: Mulching previous standing maize with wild sage followed by its incorporation with conventional tillage immediately prior to sowing of wheat	N_{60}	0.19	0.61	2271
	N_{120}	0.21	0.67	2711
T_2: As T_1 but wheat sowing following conservation tillage and retaining the residues of wild sage as mulch	N_{60}	0.26	0.68	2532
	N_{120}	0.29	0.78	2942
T_3: As T_1 but mulching by eupatorium instead of wild sage	N_{60}	0.20	0.58	1850
	N_{120}	0.24	0.65	2330
T_4: As T_2 but mulching with eupatorium instead of wild sage	N_{60}	0.25	0.60	2252
	N_{120}	0.29	0.72	2623
T_5: Farmer's practice of conventional tillage after maize harvesting	N_{60}	0.13	0.41	1231
	N_{120}	0.17	0.53	1282
LSD ($P = 0.05$) Mulch-tillage (T)		0.05	0.06	433
Nitrogen (N)		0.03	0.04	274
Interaction (T × N)		0.07	0.08	610

Adapted from Acharya CL, Kapur OC, and Dixit SP (1998) *Soil and Tillage Research* 46: 153–163.

Table 11 Tuber yield and water-use efficiency (WUE) of potato as influenced by mulching and N application (pooled over 4 years)

	Autumn potato		*Spring potato*	
Treatment[a]	*Yield (Mg ha^{-1})*	*WUE (kg ha^{-1} cm^{-1})*	*Yield (Mg ha^{-1})*	*WUE (kg ha^{-1} cm^{-1})*
60 kg N ha^{-1}	8.46	479.9	13.30	257.0
120 kg N ha^{-1}	9.86	528.2	15.15	271.2
Pine needle mulch @ 10 Mg ha^{-1} + 60 kg N ha^{-1}	12.69	711.0	16.33	340.9
Pine needle mulch @ 10 Mg ha^{-1} + 120 kg N ha^{-1}	14.52	735.5	18.34	395.4
CD ($P = 0.05$)	1.19	-	1.46	-

[a] All the treatments received a uniform dose of FYM @ 20 Mg ha^{-1}.

Reproduced from Acharya CL and Kapur OC (2001) *Indian Journal of Agricultural Sciences* 71: 306–309.

With increased water availability under mulching, crop response to fertilizer increases, thereby improving the fertilizer-use efficiency. Enrichment of soil with plant nutrients through organic mulching materials boosts crop production. Furthermore, mulching helps in moderating the soil thermal regime, which has a profound influence on nutrient transformations in soil, root growth and nutrient uptake by crop plants. Thus it enhances the nutrient-use efficiency by crops.

Limitations of Mulching

Despite the positive aspects, the following are limitations, which must be taken into account while following the practice of mulching.

1. If mulch is applied before sowing it hinders sowing and many post-sowing operations like fertilizer applications, irrigation etc. Mulch poses special problems for the arable farmer because tillage tools become clogged with the residue, thus planting under the residue is not always successful. Some specific equipment is required for post-sowing operations in mulched plots.

2. In shallow potato planting if mulch is improperly managed it may lead to a greater percentage of stolon becoming aerial shoots and more tubers becoming sunburned. The mulch layer in potato fields should be removed once full canopy is achieved, to avoid disease and pest infestation.

3. In areas of high rainfall, thick mulching may lead to water-logging conditions and provide an environment for disease infestation. If the soil is too wet under mulch, an increase in denitrifying bacteria may cause loss of N.

4. The effectiveness of a mulch depends upon the depth and extent of its coverage. However, if the mulch is applied too thickly it may hinder seedling emergence through the deep impenetrable barrier.

Therefore, it is suggested to use a thin layer of mulch at planting so as not to impede emergence and to follow with a thicker application once the seedlings are established.

5. Plastic mulches usually require pick-up and disposal at the end of the season and their manufacture and disposal entail significant environmental costs, as these are not biodegradable. Plastic mulch can interfere with the recharge of the soil profile by rain or overhead irrigation, which can lead to water deficits in vegetables if drip irrigation is not used.

6. Numerous phytotoxic substances associated with various crop residues have been shown to limit growth of several crop species reportedly through exudation, leaching or microbial production of allelopathic chemicals.

7. There are competing demands for crop residues for other uses in the semiarid tropics of developing countries. Where most of the plant residues are used as cattle fodder, farmers may not adopt mulching under conditions of scarce fodder. In the tropics, crop residues are also used for fencing, roofing and as a source of household fuel. Mulching is thus practiced under conditions where the marginal return from this practice is higher than its other competing demands.

8. The crop residue mulches reduce the temperature of the surface soil layers. In cool climates, this reduction in soil temperature shortens the growing season of the arable crops, while in wet areas higher soil moisture may induce gleying and anaerobic conditions. The limiting problem of using residue mulch on soils in temperate regions is its inability to remove excess water to allow the soil to warm up in time for planting the next crop.

9. The mulch competes with the main crop for nitrogen as it decomposes if the C:N ratio of the material is very high due to immobilization of the soil nitrogen by the microorganisms.

10. The surface residue mulch may generate a more favorable habitat for soil and surface-dwelling insects and pathogens. The residue left at the surface reduces water loss, moderates temperature extremes and provides food resources for the pests and may increase their infestation. Besides this, for many plant pathogens, residues provide a source of food, a place to live and a place to reproduce. A significant proportion of root-infecting organisms, for example, depend upon crop residues for their survival in soil. Root diseases caused by *Pythium* species are favored by wet soils and cool temperatures, especially in corn production.

See also: **Aggregation:** Physical Aspects; **Cover Crops**; **Crop-Residue Management**; **Crusts:** Structural; **Cultivation and Tillage**; **Erosion:** Water-Induced; **Evaporation of Water from Bare Soil**; **Infiltration**; **Nutrient Management**; **Water-Use Efficiency**

Further Reading

Acharya CL, Kapur OC, and Dixit SP (1998) Moisture conservation for rainfed wheat production with alternative mulches and conservation tillage in the hills of north-west India. *Soil and Tillage Research* 46: 153–163.

Bhagat RM and Acharya CL (1987) Effect of soil management on rainfed wheat in northern India. 1. Hydrothermal regime and root growth. *Soil and Tillage Research* 9: 65–77.

Bond JJ and Willis WD (1969) Soil water evaporation: Surface residue rate and placement effects. *Soil Science Society of America Proceedings* 33: 445–448.

Hatfield JL and Stewart BA (1994) Crop residue management. *Advances in Soil Science*. Boca Raton: Lewis Publishers Inc.

Lal R (1987) Farming systems. In: Lal R (ed.) *Tropical Ecology and Physical Edaphology*, pp. 618–685. New York: Wiley-Interscience Publication.

Prihar SS (1986) Fertilizer and water use efficiency through the aid of mulching. In: *Proceedings of Seminar Organized by Fertilizer Association of India* (Northern Region) at Dehradun, India, May 10–11, 1986. pp. 31–44. Dehradun: Fertilizer Association of India.

Prihar SS, Jalota SK, and Steiner JL (1996) Residue management for reducing evaporation in relation to soil type and evaporativity. *Soil Use and Management* 12: 150–157.

Stratton ML and Rechcigl JE (1998) Organic mulches, wood products and composts as soil amendments and conditioners. In: Wallace A and Terry RE (eds) *Handbook of Soil Conditioners – Substances that Enhance the Physical Properties of Soil*, pp. 43–95. New York: Marcel Dekker Inc.

Unger PW (1990) Conservation tillage systems. *Advances in Soil Science* 13: 27–68.

Unger PW (1994) *Managing Agricultural Residues*. Boca Raton: Lewis Publishers Inc.

MYCORRHIZAL FUNGI

L M Egerton-Warburton, Chicago Botanic Garden, Glencoe, IL, USA
J I Querejeta and M F Allen, University of California – Riverside, Riverside, CA, USA
S L Finkelman, Containerless Research, Inc., Evanston, IL, USA

Introduction

Mycorrhizae ('fungus roots') are an almost universal terrestrial mutualism between plant roots and certain soil fungi, the oldest fossil evidence of which is in the rhizomes of early vascular plants from approximately 400 My ago. Most mycorrhizae are mutualistic, i.e., the fungi provide soil resources to the plant in exchange for photosynthates. Six major types of mycorrhizae are recognized: ectomycorrhizae (EM), arbuscular (AM), monotropoid, arbutoid, orchid, and ericoid (Table 1; Figure 1a). These categories are based largely on differences in phylogenetic, structural, or ecophysiological differences among fungal–plant relationships. We focus on AM and EM, the most common types.

AM are unique in that they are the only monophyletic mycorrhizal lineage that has coevolved with land plants (Table 1; Figure 1b). Although there are relatively few species of AM fungi (approx. 175 species to date), they associate with a wide diversity of plants in grasslands, many tropical forests, and some temperate forests. AM are particularly efficient at the uptake and transfer of nutrients under conditions of low soil moisture or fertility, especially immobile nutrients with low diffusion rates or adsorbed to soil particles (phosphorus, P), and mobile mineral species (NH_4).

EM and the remaining mycorrhizal types demonstrate polyphyletic lineages that represent parallel or convergent evolution, interspersed with saprophytic relatives, and multiple lineages that have gained or lost the capacity to form mycorrhizae. These associations are formed by a vast array of fungal taxa (more than 5000 species; Table 1; Figure 1b). However, compared with AM associations, a mere 3% of plants are EM, most of which are woody trees and shrubs in boreal, temperate, and some tropical forests. These mycorrhizae also demonstrate the capacity to acquire N in a variety of forms (proteins, peptides, NH_4) and sources (weathered rock, leaf litter), and P by producing extracellular acid phosphatases that release P from organic complexes in the soil.

The remaining types are confined to specific plant families and/or habitats. Arbutoid mycorrhizae are a variant of EM and found in ericaceous plants. Ericoid and orchid mycorrhizae coexist with specific plant groups and harbor specialized fungal symbionts that promote plant growth by directly mineralizing organic substrates, often at low soil pH, by releasing enzymes into the substrates (Table 1; Figure 1b). Monotropoid mycorrhizae demonstrate the evolution toward increased host control over their fungal partners (mycoheterotrophy). Typically, the host plant is dependent on fungi for mineral nutrients and energy (carbon substrates), because they lack photosynthetic apparatus and/or substantial roots, but the fungi derive little or no benefit from these associations.

Apart from the construction of fungal structures on or in the root, mycorrhizae invest resources in producing hyphae, microscopic threads 2–10 μm in diameter, which grow out into the soil from the mycorrhizal root (Figure 1). These external hyphae ramify the soil and increase plant nutrient and water supply by extending the volume of soil accessible to plants, act as a sink for carbon, bind soil particles into soil aggregates, and provide food for animals. Mycorrhizal benefits can include greater plant yield and nutrient accumulation. Moreover, hyphal linkages shared by two or more plants may facilitate the transfer of resources in a community, or assist in seedling establishment.

For the most part, being mycorrhizal thus represents a considerable C expense to the plant. In community and ecosystem structure and functioning, a large proportion of plant photosynthate production is allocated to mycorrhizal roots, resulting in a large flux of C into the belowground detrital system. In fact, as mycorrhizal fungal tissue grows, between 3 and 20% of a plant's net primary productivity is transferred from the host plant and allocated to the pool of live fungal tissue. The obligate fungal requirement for host-derived C means that any environmental factor that affects plants also affects their fungal mutualists. Then again, mycorrhizal diversity also determines the competitive balance between plant species and the overall diversity and productivity of plant communities, because each fungal species may differentially enhance (or restrain) survival, growth, and nutrient acquisition among host plants. Accordingly, each change in the soil physical or chemical environment that influences mycorrhizal functioning may also affect the host plant. Such interdependencies or linkages typically characterize the central paradigm of complexity,

Table 1 Summary of the principal evolutionary and ecologic characters of the major mycorrhizal types

	Arbuscular	*Ectomycorrhizae*	*Ericoid–arbutoid–monotropoid*[a]	*Orchid*
Evolution				
Age of association (My)	>400	>100	±80	±100
Fungal–plant coevolution	+	+	*	*
Specificity	Low	Intermediate	Medium-high	High
Multiple lineages of fungi	No	Yes	Yes	Yes
Fungal taxa	Zygomycetes	Basidiomycete Ascomycete Zygomycete	Ascomycetes Basidiomycetes	Basidiomycetes
Plant taxa	Bryophytes Pteridophytes Gymnosperms Angiosperms	Gymnosperms Angiosperms	Ericales Monotropodeae Bryophytes	Orchidaceae
Ecology				
Fungi septate	−	+	+	+
Fungi aseptate	+	−	−	−
Intracellular colonization	+	+	+	+
Fungal sheath	−	+ or −	+ or −	−
Hartig net	−	+	+ or −	−
Vesicles	+ or −	−	−	−
Depends on host for survival	+	−	−	−
Host achlorophyllous[b]	− (?+)	−	− (Ericoid, arbutoid) + (Monotropoid)	+ (Seedling[b])
Plant depends on fungus for seed germination	−	−	+ (Monotropoid)	+ (Seedling)
Ecophysiology				
Nutrient transfer to plant (and nutrients transferred)	+ (P)	+ (N)	Yes (N, P)	Yes (N, P)
Carbon transfer to fungus	+	+	+ (Ericoid, arbutoid)	+ (Chlorophyllous spp.)
Carbon transfer to plant	−	−	+ (Monotropoid)	+ (Achlorophyllous spp., seedlings)
Specialized nutrient–carbon transfer interface	+ Arbuscule	+ Hartig net	− (Hyphal coils, peg)	− (Hyphal coils)

Character states: +, present; −, absent; *, not known.

[a] Members of the Ericales are host to arbutoid, ericoid, and monotropoid mycorrhizae. For clarity, these types are considered as one group.

[b] All orchids are achlorophyllous in the early seedling stages; however, the extent of chlorophyll development varies among adults.

from the Latin *complexus*, meaning 'entwined' or 'twisted together.'

Complexity

What is complexity? What does it mean to have a complex mycorrhizal system? Traditionally, mycorrhizal functioning has been evaluated by empirical and descriptive studies or by theoretical concepts, mathematical models, and deductive structures, i.e., analytic versus synthetic models. Most experiments are also reductionist. In other words, they successively reduce a mycorrhizal system into smaller and simpler components that are studied in isolation and in a formulated manner. In contrast, the complexity approach is rich and holistic. The individual components within a mycorrhizal system mean little on their own, but together they give rise to the collective responses or emergent patterns of the whole system, and how this system interacts with its environment (Figure 2). This tenet is also central to ecology, or the totality of relations between organisms and their environment.

Mathematically, such systems exist 'on the edge of chaos,' or somewhere between perfect order and complete randomness, where a small change in one component can lead to widely varying results overall. Moreover, complexity *per se* has been allied with terms such as 'fractals,' 'cellular automata,' 'genetic algorithms,' 'order and hierarchy,' to name just a few. Such terms, however, simply emphasize the many different approaches that may be used to understand complexity. Can one use the science of complexity to understand mycorrhizal systems better? Conventional approaches based on linear causality and stochastic properties, while illustrative of the individual components, are inappropriate, because by their very nature mycorrhizal systems are dynamic processes rather than static entities. The canon is complex interplay. Conceptual models might provide

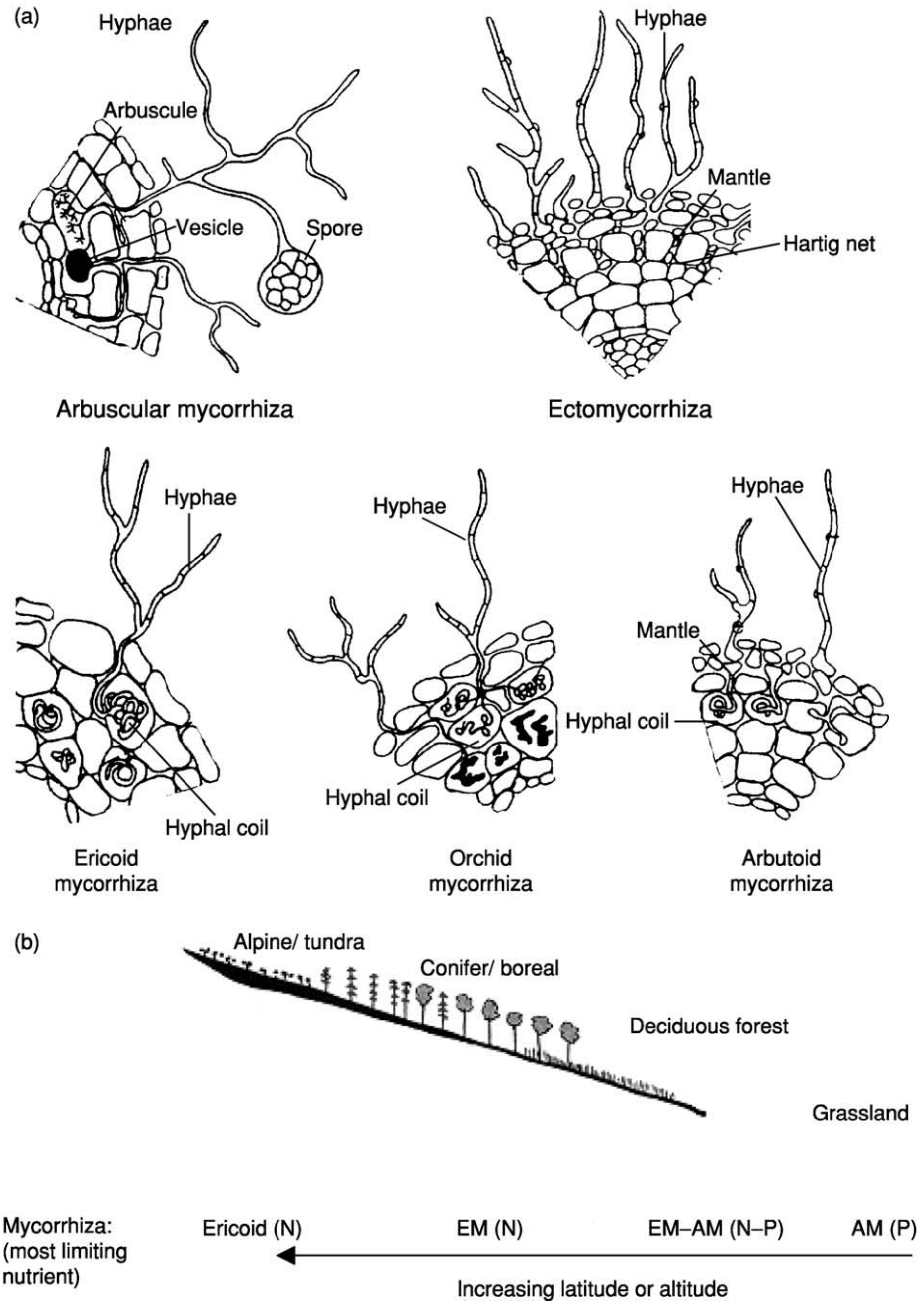


Figure 1 (a) Differential relationships between fungal and plant structures for each of the main types of mycorrhizal associations. Reprinted from Selosse *et al.* (1998) *Trends in Ecology and Evolution*, vol. 13: 15–20, with permission from Elsevier. (b) Mycorrhizal types as related to landscape position. Adapted from University of Wisconsin, Forest Ecology and Management, Department Literature.

frameworks for the identification of processes, structures, and relationships within complex systems.

Defining the Mycorrhizal System and Participants

The system is a specified region in space or a dynamic surface that changes over time (Figures 2 and 3). For example, the oak woodland bounded by grasslands and the hyphal network can both be considered as mycorrhizal systems. They are persistent and respond to the environment or repeated instances of similar or related conditions. Mycorrhizal systems repeatedly demonstrate a negative response to N fertilization and a positive response to elevated atmospheric CO_2. The environment includes organic and mineral nutrients, water, carbon, and influences such as urbanization and global change.

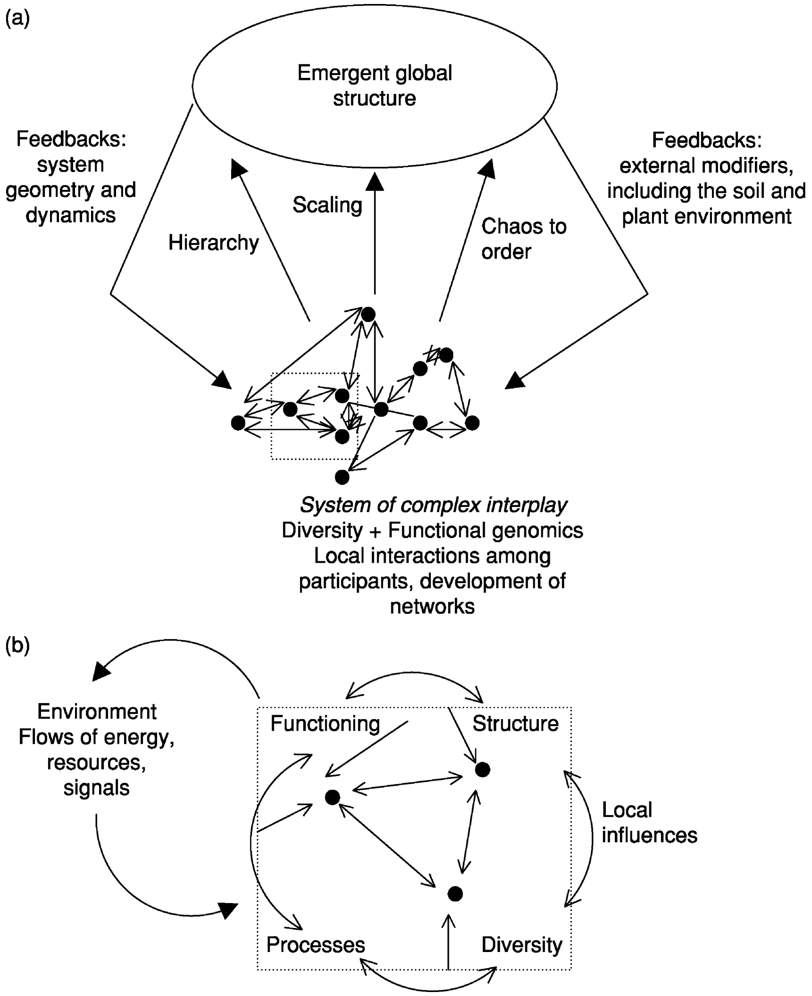


Figure 2 (a) Conceptual model of complexity as a framework for developing analyses in mycorrhizal systems; (b) components of the interactions between the mycorrhizal systems and the environment.

The system is composed of fungal and plant participants. The fungal participants may consist of single-species situations or a set of interacting species. The plant participants are networked into the system by fungi such that the system consists of a lattice (hyphal network) with a discrete agent (host plant) at each lattice site or node (Figure 3a,b). In the field, each mycorrhizal fungus attaches to as many plants as it encounters. For example, there is considerable overlap of EM associated with Ponderosa pine, Douglas fir, Tan oak, Grand fir, and Madrone. Every plant within the network interacts because any transition or change in one fungal partner leads to changes in the neighboring or connected plants. Adding new individuals progressively alters the size and degrees of connectedness within the network in a geometric fashion (Figure 3b). Connections among networks can result in increasingly complex configurations (Figure 3c). The journey into mycorrhizal complexity starts with ecologic diversity, and the structural and functional genomics of the symbiosis.

Governance of the System

Diversity and the Delivery of Ecosystem Services

The potential global taxon richness is extremely high for both plants and fungi forming mycorrhizal systems. The measures of α, β, and γ relate the number of species to area: α is the number of species per hectare (local), β, the number of species per 10–1000 ha (regional) and γ, the number of species in more than 10 000 ha (landscape). Both α and β diversity are similar among habitats and appear to reach upper limits of 5–20 species ha^{-1} for α diversity and 30–50 species for β diversity. Hence, there may be some limit to the system services provided by mycorrhizal fungi. At the γ level, patterns of diversity

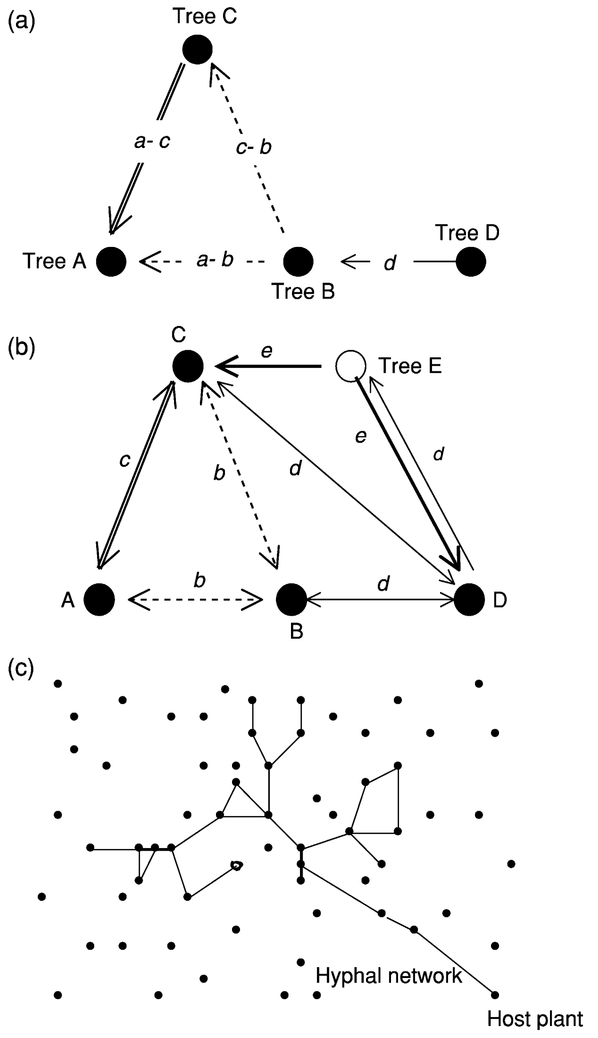


Figure 3 (a) The simplest configuration of a network among three trees; (b) the addition of another host tree and associated fungi increases the number of potential links in the network; (c) the addition of subsequent host plants further increases the complexity of links among plants and the system as a whole. Upper-case letters and filled circles denote the host tree; lower-case letters denote fungus or fungi associated with a host tree. Arrows denote the potential hyphal linkages among plants.

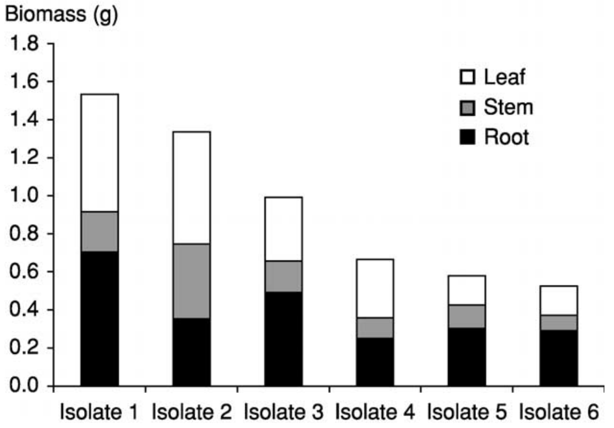


Figure 4 Biomass accumulation in the leaves, stems, and roots of *Quercus* plants grown with six different cultures of *Cortinarius* that were isolated from one clump of sporocarps.

among mycorrhizal types diverge. Forty-eight species of AM fungi have been recovered from *Artemisia tridentata*, although only 175 species have been described globally. In contrast, more than 2000 EM species associate with Douglas fir, but worldwide the number of described EM species approaches many thousands.

What is being measured? Molecular approaches measure species and population genetic diversity, whereas sporocarp (mushroom) and morphotype estimates are responsive to host species or environment. Thus, sporocarp diversity is unlikely to reflect the diversity of EM fungi on the root tips. In *Quercus* woodlands, only three species of EM sporocarps per tree have been collected compared with seven to nine morphotypes and up to 20 molecular types per tree. In addition, a few widespread species of fungi account for more than 50% of mycorrhizal abundance and the majority of species are rarely detected. In Douglas fir, 69 morphotypes of EM have been recovered, of which nine types colonize 67% of the root tips.

Fungi also separate both spatially and temporally even when associated with the same host plant. Spatial variation within a system is high because the dominant fungi vary considerably among trees. Fungi may also be spatially segregated to the undercanopy, edge of canopy, and interspace areas. From the temporal perspective, EM formed by certain species of *Laccaria*, *Amanita*, and *Hebeloma* persist in small genets for an average of 1–2 years. In contrast, large genets that last decades are formed by species of *Suillus* and *Cortinarius*. It follows that all species have the potential to contribute to system functioning, although not necessarily in synchrony.

Increasing the diversity of fungi within a system can increase the diversity, productivity, and stability of the plant community. Douglas fir grows better when supported by multiple mycorrhizal fungi, and especially in competition with Ponderosa pine. Different fungi can also undertake different ecologic roles across space and time. For example, six isolates of *Cortinarius* isolated from a single group of sporocarps produce different biomass allocation patterns in their host plant (Figure 4). Certain fungi may preferentially acquire organic nitrogen (N) (*Boletus*) and others, inorganic N (*Pisolithus*, *Laccaria*). *Cenococcum geophilum* exploits the local soil for resources, whereas *Pisolithus* forms extensive hyphal networks that run from one resource patch to another. *Rhizopogon subcaerulescens* persists deep within the soil profile, survives fire, and thus forms mutualisms in the postfire environment, whereas the upper-profile dweller, *Thelephora*, does not. Such patterns of diversity have important implications for system functioning.

Multiple Genes Coordinate the Development of Mycorrhizae

Both the fungal and plant partner engage in a precisely orchestrated developmental sequence that results in the expression of distinct fungal or plant genes exclusively or predominantly within the symbiosis. In *Eucalyptus globulus–Pisolithus tinctorius* EM, 65 genes record the dialogue of the early development of the fungal mantle. Mycorrhizal development results in the upregulation of the translation initiation factor eIF4A gene, a high-molecular-weight (MW) protein complex, in *Pisolithus tinctorius*, sequences coding for classes of symbiosis-regulated (SR) genes, particularly the cysteine-rich hydrophobins and multigene SRAP32 families, and more recently proline-rich proteins and proteophosphoglycans. eIF4A is essential for translation and cell growth, whereas the structural proteins encoded by SR genes are usually involved in membrane synthesis and function. Similarities to genes encoding proteins involved in primary metabolism and transport (carbon, H^+-ATPase), the cell cycle, deoxyribonucleic acid (DNA) replication, cell structure (β-tubulin), fungal structure (arbuscule degradation, hyphal proliferation), signal transduction, and the suppression of plant defense genes have also been recovered from AM. In addition, mutants unable to sustain AM (Myc^-) were also unable to associate with N-fixing Rhizobium (Nod^-). Conserved signal transduction pathways may thus be common to both mutualisms during their early development

Genomic Hegemony over Nutrient and Carbon Transport

Just as many genes can be ascribed to development of the symbiosis, so can the expression of other functional genes record the dialogue of carbon and nutrient transfer in mycorrhizae. For example, the monosaccharide uptake system in *Amanita muscaria* requires a transporter, AmMST1. AmMST1 contains the open reading frame for a 56-kDa protein that is expressed when *A. muscaria* is axenically cultured at a glucose concentration of less than $5\,\text{mmol}\,\text{l}^{-1}$(the threshold expression level) for 18–24 h, and in *A. muscaria–Picea abies* and *A. muscaria–Populus tremula* × *tremuloides* mycorrhizae. The expression of AmMST1 and any subsequent increase in the import of monosaccharides by fungus triggers an increase in the supply of photoassimilates from the host. Phenylalanine ammonia lyase (AmPAL), a key enzyme of secondary metabolism (phenolics), is also involved in plant defense against pathogenic microbes. Phosphoenolpyruvate carboxylase (PEPC) is expressed in regions of intense mycorrhizal development; PEPC moderates the assimilation of NH_4 in fungi and the transport of amino acids such as glutamine from fungus to plant. However, many genes in fully sequenced genomes are of unknown function, because mycorrhizae are typically studied in a synthetic environment and outside the ecologic context in which they evolved.

Functional Genomics

Responses of a mycorrhizal system to specific environmental cues are tightly coordinated by a physiologically appropriate subset of its genes. Nutrient transport from soil to plant via the mycorrhizal fungus has been the hallmark of mycorrhizal studies for more than a century. One gene that may be involved in P transport efficiency of AM is a fungal membrane P transporter, which moves P from the soil into the hyphae. This gene is regulated by the availability of P in the soil environment, but the fungal partner controls the synthesis of polyphosphate (polyP) and thus the amount of P available in the root and shoot. In turn, *in vivo* ^{31}P nuclear magnetic resonance (NMR) illustrates that mycorrhizal plants typically accumulate P as supramolecular aggregates of polyP (chain length 5–11 units), whereas nonmycorrhizal plants accumulate solely inorganic P (P_i). In mycorrhizae, polyP represents much of the mobile P reservoir.

Mycorrhizal fungi are also capable of utilizing many forms of N and along a variety of pathways. A putative nitrate-transporter gene has been isolated from mycorrhizal *Medicago* in which the expression is downregulated by some, but not all, AM. NMR reveals that NH_4 is predominantly incorporated by glutamine synthetase/glutamate synthetase pathway (GS/GOGAT), with a small proportion by the glutamate dehydrogenase (GDH) pathway. AM may also accelerate the decomposition of organic matter and acquire N directly from these substrates. ^{15}N pulse-chase studies show the incorporation of labeled N into arginine, which is consistent with the ornithine cycle. In turn, the N is stored in relatively immobile complexes with polyP in vacuoles.

Fractionation of N during N uptake by fungi results in considerable shifts in ^{15}N abundance such that ^{15}N abundance in fungal tissues is unlikely to reflect those of the original soil N. The patterns of fractionation are also fungus and N source-specific. Intense fractionation against ^{15}N in ericoid systems results in a ^{15}N-enriched plant, whereas EM systems discriminate against ^{14}N (^{15}N-depleted plant). In addition, δ^{15}N values for EM sporocarps in a *Quercus* woodland range from 2 to 11. Thus, organic and inorganic N compounds can be accessed differentially by fungi. ^{15}N natural abundance also demonstrates that the strict functional boundaries between mycorrhizal and decomposer fungi may be less clear-cut than previously considered (δ^{15}N 5.29% in mycorrhizal *Boletus*; 6.03% in saprobic *Bovista*).

Spectra from ^{1}H- and ^{13}C-NMR assays in AM *Gigaspora* demonstrate that fungal spore germ tubes, extraradical hyphae, and intraradical hyphae differ in their lipid (triaglycerol) and carbohydrate (hexose) metabolism. Spore germ tubes are sites of lipid breakdown and gluconeogenesis. In intraradical hyphae, hexose is transported into the cytoplasm, where lipid synthesis occurs. Extraradical hyphae act as sink for lipids, and the lipids are converted to carbohydrates (trehalose) and occasionally glycogen, and stored in spores or vesicles. Fungi convert the simple sugars to trehalose or other complex sugar to prevent or reduce reabsorption. Changes in root phytohormone composition also occur in tandem with the loading and unloading of C.

Nutrient fluxes may also be coregulated because the translocation of nonlimiting nutrients such as magnesium (Mg) depends on simultaneous long-distance translocation of P. Combined mass balance (P, potassium, K) and stable isotope analyses (^{25}Mg) show that hyphae can transfer up to 5% K and 3–4% Mg as long as P fluxes through the system.

The factors regulating the amounts of resources transported between fungi are poorly understood. Virtually all studies to date have concentrated on demonstrating the presence and form of elemental transport from soil to plant via the fungal hyphae. As a result, there is little quantitative work on the mechanisms regulating the transport of resources among fungi.

Games and Models for Mapping Causality

Knowing the structure of water yields no clues as to why water goes down a drain in a vortex. Similarly, in mycorrhizae, the genes are expressed and the metabolic machinery produces parts of the system, but they provide no information on how structure in mycorrhizal systems is generated. Mapping is a way of generalizing dynamics and establishing conceptual illustrations and models using simple input/output diagrams. The cognitive system map (Figure 5) draws a causal relationship among components and predicts how complex events might play out. Each arrow defines a causal link where 'plus' is an increase or enhanced response, 'minus' is a decrease,

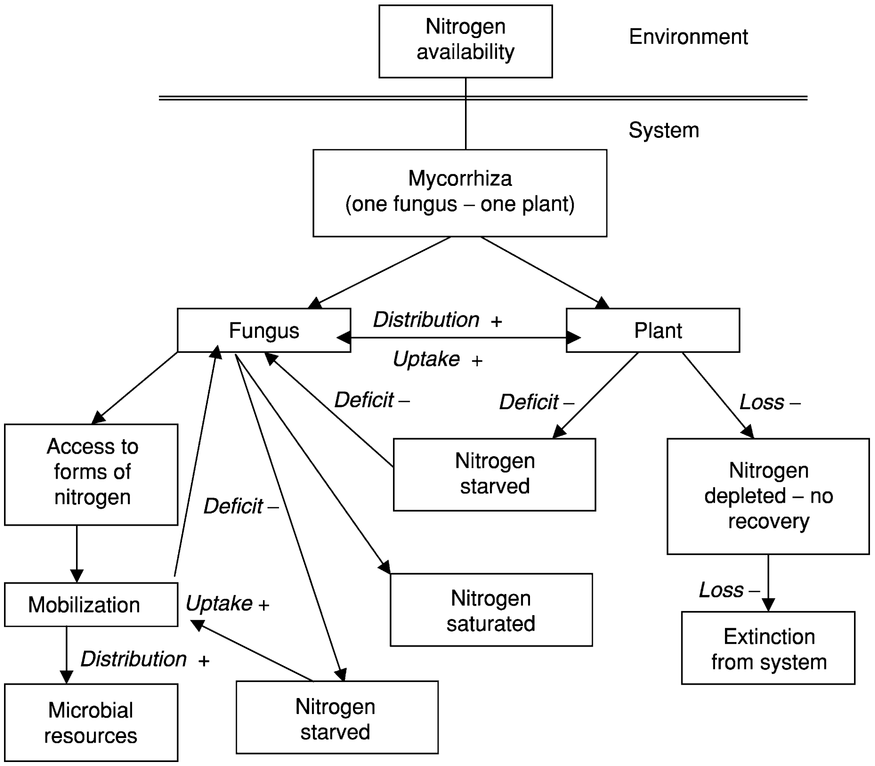


Figure 5 A single-state model of the relationship between the environment and the mycorrhizal system in response to soil nitrogen status.

or negative response, and 'zero' is stasis. The final cognitive map defines the nonlinear dynamic system.

This single-state map for a one plant–one fungus mycorrhizal system and its response to soil N availability illustrates that the fungal response is enhanced by plant nutrient stress and decreased by an increase in nutrient availability and drought. The plant response is enhanced by P availability and decreased by drought. The overall model is the 'If condition, then action,' which has the same meaning as statements in programming language. If the condition portion is satisfied, then the action is triggered and other rules are activated. This is the 'bucket brigade' algorithm (*sensu* J.H. Holland), where signals are passed down the line to reinforce a chain of effect. This is also similar to path analysis. Calculations of the potential outcomes (fungus, plant, system, or 3^n) from the model (Figure 5), indicate that if soil N is limited, then the fungus transports approximately 64% of N to plant, or strengthens the importance of the fungus to the system. However, if soil N increases then the fungus transports only approximately 34% of plant N.

Adding more participants into the system increases the number of possible outcomes exponentially. A two-fungus system responds quite differently to the environment than a one-fungus system. Fungal respiration is depressed in each of *Rhizopogon* and *Suillus* mycorrhizae by 72% and 30%, respectively, after the addition of NH_4 (Figure 6). An additive cognitive map, however, indicates that NH_4 may have a bigger negative effect on respiration in *Rhizopogon–Suillus* in concert than on a single species or the edges where they meet. Using mapping rules as patches also demonstrates the nonlinear nature of the sporocarp $\delta^{15}N:\delta^{13}C$ divide between mycorrhizal and saprobic fungi (Figure 7). In particular, when $\delta^{15}N$ is ± 0, $\delta^{13}C$ in mycorrhizal fungi is depleted, whereas $\delta^{13}C$ in saprobes is enriched.

As more knowledge emerges, so do more rules and a more elaborate model that is connected to actual observations. For example, the N needs of the fungus and plant are determined by the level of internal N reservoirs that are depleted at a rate that keeps the fungus and plant functioning, so that low N implies a high need for the resource. The N level in the reservoir may only be increased by providing the correct signal molecule or transporter. Adding these new parameters into Figure 6 increases the number of potential iterations and outcomes and, in turn, the complexity of the system. Such complex systems are controlled by feedback mechanisms between plants and their mycorrhizal mutualists and soil microbes, and local host-dependent preferences for certain communities of fungi.

Hierarchy and Scaling

Complex adaptive systems are hierarchically nested because they are large systems that are comprised of progressively smaller systems, etc.; these large

	Respiration (μg C mg^{-1} ergosterol h^{-1})		
Species	*Control*	*NH_4*	*% decrease*
Rhizopogon sp.	5.1	1.4	72
Suillus pungens	3.0	2.1	30

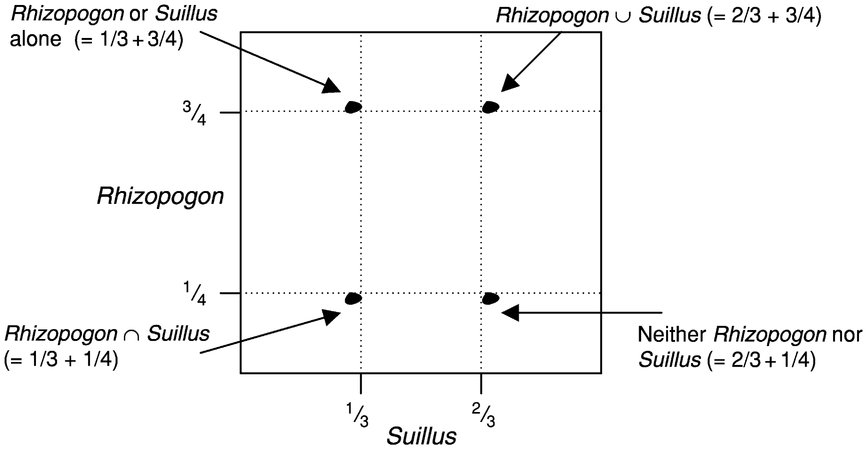


Figure 6 An additive map of the predicted interactions between the ectomycorrhizal fungi, *Rhizopogon* and *Suillus*, after media enrichment with ammonia. Ranking of analysis (highest to lowest effect) shows that NH_4 influences: two species together in culture (*Rhizopogon* ∪ *Suillus*, 17/12); single-species cultures (*Rhizopogon* or *Suillus* alone, 13/12); fungal contaminants (neither *Rhizopogon* nor *Suillus*, 11/12); patches where each species intersects (*Rhizopogon* ∩ *Suillus*, 7/12).

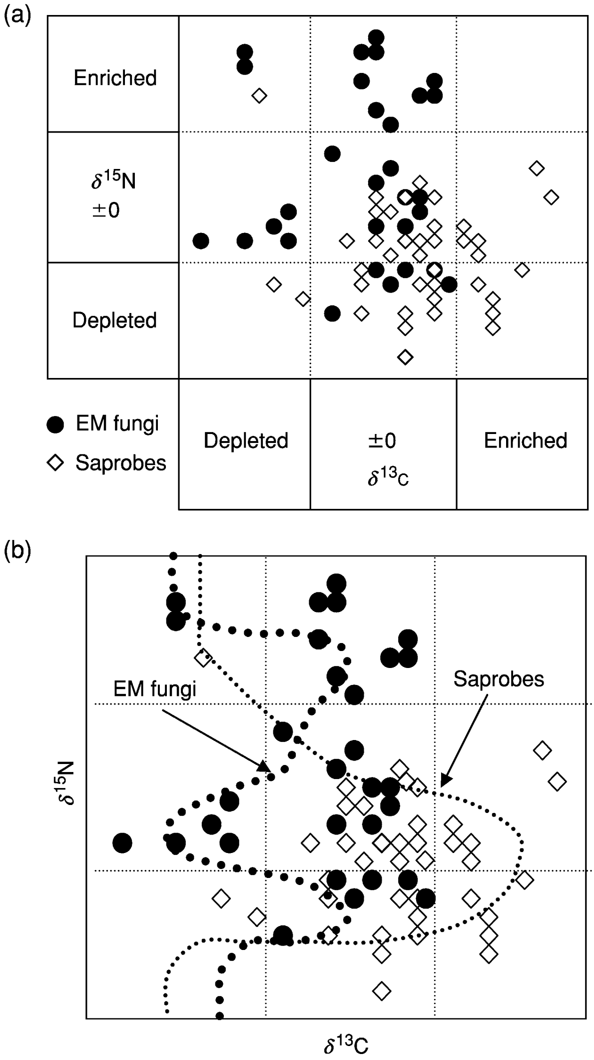


Figure 7 (a) Ratio of $\delta^{15}N:\delta^{13}C$ in sporocarps of mycorrhizal and nonmycorrhizal (saprobic) fungi; (b) additive patch map of $\delta^{15}N:\delta^{13}C$ in mycorrhizal and nonmycorrhizal fungi. EM, ectomycorrhizae.

systems can interact with each other as well. The issue of scaling and translating information from small-scale patterns and processes to understanding of large-scale problems, such as global change, is essential. This is captured in a simple way in the study of chaos and fractals. Chaos is a rich description of a complex nonlinear system that can include the coupling of different scales over time or space and where a small, seemingly insignificant change in one of the attributes of a single agent can have a widely varying effect on the system as a whole. A fractal is a noninteger semidimension, and one of its traits can be self-similarity at different scales.

Quantifying global species' diversity from measurements of local diversity (species, transcripts, sequences) is typically summarized as a function of sample size and area, i.e., a species–area curve (Figure 8). The classic species–area curve asymptotically approaches the total diversity in the collection. On the contrary, species–area curves for EM morphotypes typically proceed in a step-wise fashion and bear a stronger resemblance to the Cantor (or Devil's staircase) fractal. An inverse relationship exists between species diversity and *D*, the dimension of the Cantor function. Therefore, systems with high diversity have low *D* (dim → <0), and conversely, mycorrhizal systems with low diversity possess a high *D* (dim → <1). Relating *D* to issues in global-change biology suggests that an increasing input of N may correspond to a progressive increase in *D*, while elevated atmospheric CO_2 may correspond to lower *D*.

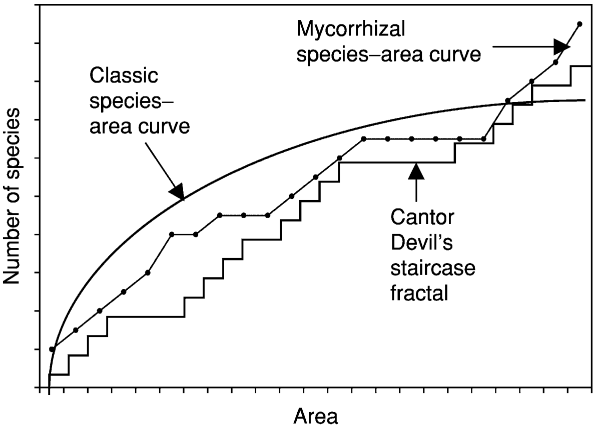


Figure 8 Overlap of the asymptotic (classic) species–area curve, the species–area relationship for morphotypes in oak woodlands, and the Devil's staircase fractal.

Dimensionality within the Cantor function also appears to corresponds with the transition from order to chaos, because entropy (disorder) and enthalpy (energy) are inextricably linked to chaos as follows: maximum diversity: dim → 0: minimum order: maximum disorder: minimum energy. This follows from Shannon–Wiener's adaptation of Newton's second law of thermodynamics as applied to information theory. A commonly used measure of diversity, the Shannon–Wiener index, is also calculated using exactly the same formula as for statistical entropy. Thus, calculations of diversity are simply reflections of system order or disorder (chaos), where the higher the diversity index the greater the level of chaos within a system.

Autonomous Networks

Fungal and plant species differ in mutualistic functioning and tolerance to the environment, but, when networked, the benefits of each are distributed

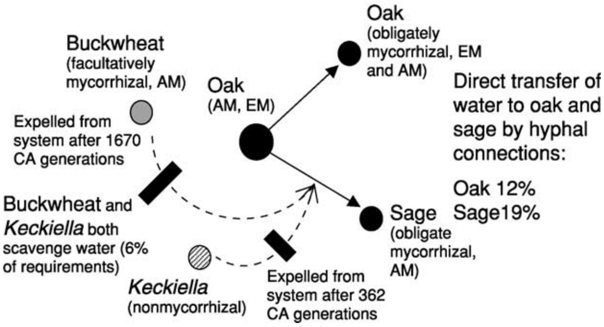


Figure 9 Model estimates of the time and sequence of plant species elimination from a fungal network based on simple cellular automata (CA) rules of water sharing and host mycorrhizal status. AM, arbuscular mycorrhizae; EM, ectomycorrhizae.

throughout the system. Fungal networks allow both plants and fungi to share resources, including carbon and water (Figure 9).

Network models based on cellular automata (CA) are appropriate in nonlinear systems. Traits or variables such as diversity (entropy) may be used for CA and, in turn, a set of cellular automata may form a Cantor set. In the simplest case, a CA consists of fungal hyphae that evolve in a series of time steps and according to simple specific rules. Even though the rules may be very simple in the initial state, the pattern produced is self-similar (or fractal). Fractal analysis derives the models of hyphal growth. EM hyphae demonstrate branching in a binary tree in which each of the nodes has two branches and operations are restricted to a two-way choice (left or right, great or less). *L*-systems describe system growth and efficiency. Because EM fungi follow the Twig model and AM fungi follow Weed models, the *L*-system can be used to describe mathematically or simulate mycorrhizal growth patterns. This is packing efficiency, or the manner in which mycorrhizal systems can squeeze a great deal of hyphal length and surface area into a finite volume and, in turn, resource acquisition. Because the production of extraradical hyphae differs significantly among mycorrhizae, packing efficiency may be a useful indicator of system functioning.

Integrating this knowledge into mycorrhizal functioning indicates how the network might proceed. Plants joined by a hyphal network may share water by either direct or indirect transfer (Figure 9). Using a CA with simple water-sharing rules, nonmycorrhizal *Keckiella* is eliminated from the group after 362 generations and facultative mycorrhizal *Eriogonum* after 1670 generations. Only plants that are directly connected by fungal hyphae remain in the system, because packing efficiency (resource transfer) increases within the networked area.

The extent of network development and the transfer resources among plants is still an open question. Spanning trees may be useful in calculating these outcomes. In addition, multiple transfers of resources suggest redundancy. Different fungal species, and possibly communities, may be capable of the same processes even though there may be different versions of the gene(s) that encode for a certain purpose, e.g., N acquisition.

Summary

Lovelock's Gaia hypothesis conceptualized biodiversity and mutualism in their most advanced and elegant integration. In mycorrhizae, diversity and mutualistic functioning unite successive systems into networks and complex systems. In order to show that complexity has increased overall, it is sufficient to show, that – all other things being equal – connections have increased in at least one dimension. What is lacking is the ability to make predictions about how complexity in mycorrhizal communities will change their function as the systems and/or their environment is altered by human impacts and global change. Combining molecular, metabolic, and biochemical data with nonlinear mathematical models might provide the foundations and rules for understanding mycorrhizal complexity. The limitations and utility of any data, however, remain in developing data-mining and complexity-modeling tools and techniques to utilize effectively the information from a local and global perspective, because data are gathered on scales from molecules to genomes, organelles, cells, tissues, and organs. Bioinformatics is the acquisition of knowledge by means of computational tools for the organization, management, and mining of genetic biological data. These analytical tools are generally applied to the oceans of data collected by molecular biology, e.g., the Human Genome project. A more appropriate term for mycorrhizal systems may be 'ecoinformatics' or the accumulation of ecologically based data sets appropriate to mycorrhizae *in situ*, followed by data integration. In doing so, it will then be appropriate to say that diversity and mutualism provide ecosystem function and what that functioning may be.

See also: **Fungi**; **Root Architecture and Growth**; **Root Exudates and Microorganisms**

Further Reading

Allen MF (1996) *The Ecology of Mycorrhizae*. Cambridge, UK: Cambridge University Press.

Smith SE and Read DJ (1997) *Mycorrhizal Symbiosis*, 2nd edn. San Diego, CA: Academic Press.

Van der Heijden MGA and Sanders IR (eds) (2002) *Mycorrhizal Ecology*. Berlin, Germany: Springer-Verlag.